UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

A discrete, logical modelling framework to study tissue patterning and morphogenesis

Camila Veludo Ramos

MESTRADO EM BIOINFORMÁTICA E BIOLOGIA COMPUTACIONAL
ESPECIALIZACÃO EM BIOLOGIA COMPUTACIONAL

Dissertação orientada por:
Prof. Dr. Claudine Chaouiya
Prof. Dr. Francisco Dionísio

2016

Resumo

A formação de padrões é um processo espacial e temporalmente complexo. Os mecanismos subjacentes envolvem redes regulatórias intracelulares sensíveis a sinais microambientais, dos quais se destaca a comunicação intercelular. Modelos matemáticos fornecem um grau de abstracção conveniente para integrar e testar o conhecimento de tais mecanismos, associando as redes celulares à dinâmica populacional. Neste contexto, abordagens discretas apresentam-se como adequadas para modelar redes regulatórias de grandes dimensões que podem conter centenas de componentes.

Neste trabalho introduzimos um quadro discreto adaptado à modelação da formação de padrões epiteliais, implementado numa ferramenta gratuitamente disponível, o EpiLog. Este quadro consiste num autómato celular bidimensional de células hexagonais cujo comportamento é modelado por Modelos Lógicos a elas associados (Booleanos ou multi-valores).

O primeiro tópico desta dissertação consiste na demonstração da introdução de estocasticidade que permite superar a sincronia inerente a automatos celulares, levando à formação de padrões complexos a partir de populações equipotentes de células indiferenciadas.

O nosso principal foco de atenção é um recente método que possibilita a inclusão de proliferação celular no autómato, permitindo o estudo do crescimento epitelial. A divisão celular é controlada pelo Modelo Lógico, associado às células. Após a divisão de uma célula, um novo elemento surge numa posição adjacente à original, o que requer uma reorganização do tecido. Mantendo um nível de abstracção semelhante, desprezando conformações das células, restrições físicas ou movimentos direccionados, apresentamos um método que define a

posição da célula filha adjacente, e o deslocamento das células vizinhas. Este método está assente sobre uma medida simples que define a compressão exercida sobre uma célula em função do número de células vizinhas a várias distâncias. Variando a contribuição das células vizinhas na medida de compressão, as nossas simulações apresentam diferenças qualitativas significativas na forma dos epitélios em crescimento. Adicionalmente, tendo em conta a possível regulação da actividade mitótica por efeitos de densidade local, associámos uma medida de densidade aos Modelos Lógicos, o que causa grande impacto na forma dos tecidos.

Estas extensões foram aplicadas ao desenvolvimento de redes celulares responsáveis pela formação de padrões de mecanorreceptores na superfície corporal de insectos da ordem *Diptera*, expondo o potencial deste quadro, bem como as suas limitações.

Este trabalho demonstra as vantagens no uso de abordagens discretas na modelação, para investigar processos de desenvolvimento dependentes de regulação complexa a nível intra e inter-celular.

Palavras-chave: Autómatos celulares, Redes regulatórias, Sistemas dinâmicos discretos, Formação de padrões, Diferenciaçâo celular

Abstract

Pattern formation is a complex, time and space-specific process. Underlying mechanisms involve intra-cellular regulatory networks responding to microenvironmental cues among which cell-cell communication plays an important role. Mathematical modelling is convenient to integrate and test current knowledge of such mechanisms, considering both intra-cellular networks and cell population dynamics. In this context, discrete frameworks are well-suited to model large interacting networks comprising several hundreds components.

We introduce a discrete modelling framework for pattern formation in simple epithelia implemented in a freely available tool, EpiLog. The framework consists in a cellular-automaton defined as a 2D grid of hexagonal cells whose states are driven by their associated logical (Boolean or multi-valued) models.

Here, we demonstrate how the introduction of stochasticity allows overcoming the issue of synchrony in such cellular automata, enabling the formation of complex patterns from equipotent groups of *naive* cells.

Our main focus is a recent extension to account for cell proliferation, *i.e.* growing epithelia. Cell division is directed by the intra-cellular model, thus depending on the cell state. Upon division, the daughter cell occupies one neighbouring position in the grid, requiring a reorganisation of the whole. Keeping a similar high level of abstraction, that is, not accounting for cell shape, physical constraints or directional movements, we propose a method to compute the daughter cell's position and the displacements of the other cells. This method relies on a simple measure defining how surrounding cells at different distances contribute to the compression experienced by a cell. By

varying these contributions, simulations show significant changes in the shape of the growing epithelium. Furthermore, considering that local density controls cell division, a density measure is associated to the intra-cellular network. Simulations demonstrate the impact of the density values on the epithelium shape.

The extensions have been applied to defining networks models regulating patterns of bristles in *Diptera* body surfaces, exposing the framework's capabilities and current drawbacks.

This work supports the use of discrete modelling to investigate developmental processes highly regulated at both cellular and inter-cellular levels.

Keywords: Cellular automata, Regulatory networks, Discrete dynamical systems, Pattern formation, Cell fate decision

Resumo Alargado

O desenvolvimento de um organismo depende de mecanismos regulatórios que permitem a especificação correcta de linhagens celulares que dão origem às diferentes partes do corpo.

Esta especificação apresenta-se sob a forma de padrões complexos, nos quais as células envolventes comunicam entre si, respondendo aos sinais microambientais através de mecanismos de transdução que compreendem redes regulatórias contendo centenas de componentes.

As redes celulares devem parte da sua robustez à complexidade combinatória de interacções entre moléculas, que formam circuitos de retroacção interligados.

A modelação de redes lógicas considera os componentes celulares como entidades discretas, às quais se associam valores qualitativos de expressão (no caso Booleano, 0 ou 1 para ausência ou presença do componente). Entre eles, existem interacções direccionadas, que definem o estado de um componente regulado como uma função lógica do valor de expressão dos seus regulatores. Os componentes que não possuem nenhum regulador são as entradas do sistema. A dinâmica destes modelos é dada por transições entre estados, descrevendo os valores dos componentes do sistema a cada iteração.

Este tipo de modelos é direccionado ao estudo de diferenciação celular a longo prazo, onde a especificação de um destino celular é associada à expressão combinada de vários marcadores moleculares, sem que seja necessário o conhecimento cinético inerente à regulação.

Adicionalmente, perturbações podem ser aplicadas a modelos lógicos. Estas representam alterações dos componentes da rede e permitem enunciar previsões úteis. Por exemplo, podem indicar os possíveis

alvos das redes que permitem recuperar fenótipos normais a partir de estados alterados.

A definição de uma abordagem que integre o conhecimento da regulação intra-celular ao nível epitelial permite o estudo sistemático das interacções que levam à formação de padrões e regulação subjacente.

Neste contexto, os autómatos celulares constituem uma abstracção apropriada para modelar dinâmicas populacionais. Um autómato celular consiste numa matriz finita na qual cada entrada está associada a uma célula ou indivíduo da população. Adicionalmente, um conjunto de estados possíveis é atribuido às células, representando possíveis diferenciações. O estado das células é definido em função dos seus estados anteriores, e dos estados dos vizinhos locais, formulando a comunicação entre células.

Os automatos celulares simulam a formação de padrões complexos a partir das interacções locais entre as células, que modulam os seus estados de diferenciação. No entanto, o número de estados definidos é geralmente reduzido, pois o seu efeito combinado na sinalização a uma célula por parte dos seus vizinhos é impraticável para um grande número de estados.

O EpiLog é uma ferramenta gratuitamente disponível que implementa um autómato celular bi-dimensional de células hexagonais, para a modelação da formação de padrões em epitélios simples. As células deste autómato contêm modelos lógicos de redes celulares como acima descritos. Deste modo, o estado de cada célula é dado pelo estado do seu modelo lógico. Adicionalmente, a comunicação entre as células do autómato é definida como uma regulação sobre as entradas dos seus modelos lógicos. Esta regulação depende de sinais emitidos por células vizinhas, ou seja, componentes internos das suas redes lógicas, o seu nível de expressão, e a distância à qual se conseguem propagar.

Assim, a definição de modelos epiteliais pode ser directamente comparada com resultados experimentais, sem que seja necessário conden-

sar interacções célula-a-célula, como acontece nos autómatos celulares clássicos.

Este trabalho apresenta dois objectivos principais, que conduzem a formalização e implementação de métodos que auxiliem à correcta definição de modelos de formação de padrões em epitélios.

A primeira parte da dissertação tem por base a obsevação de que os sistemas biológicos são sujeitos a ruído espacial e temporal, que contribui para a criação de ordem a partir de um estado inicial indiferenciado. Os autómatos celulares são inerentemente síncronos, no sentido em que todos os elementos são actualizados simultaneamente, com respeito ao seu estado anterior e ao estado anterior dos seus vizinhos.

Na ausência de ruído, é frequente a presença de comportamentos cíclicos sem significado biológico. Deste modo, a indução de assincronismo fornece um método para ultrapassar essa limitação. O assincronismo temporal restringe a actualização dos estados a subconjuntos dos elementos da matriz a cada iteração. O assincronismo espacial reflecte as ocasionais falhas de comunicação que podem ocorrer entre as células de um tecido, passando informação a uma célula que não corresponde à informação do real estado do sistema. Juntos, estes dois métodos permitem a geração de padrões de Turing a partir de mechanismos de activação-inibição. Enquanto que o principal responsável pela disrupção da sincronia é o assincronismo temporal, o assincronismo espacial contribui para a uniformização dos padrões entre simulações estocásticas (para o mesmo conjunto de parametros).

O segundo tema deste trabalho foca-se na formalização de um método que permita a integração de dinâmicas epiteliais (crescimento e morfogénese) no actual quadro de modelação. De forma a manter o mesmo nível de abstracção, a mitose é definida como sendo o resultado de um conjunto de estados dos modelos lógicos das células. Quando uma célula mãe se divide, duas células novas surgem: uma na posição da célula original, e outra numa posição adjacente. Na ausência de

informação sobre restições físicas ou forma e volume das células, o reajustamento do tecido a este evento dá-se por deslocamento das células. A direcção do deslocamento é definida por uma medida de compressão aplicada sobre as células, que depende do número total dos seus vizinhos. Consequentemente, num tecido onde a proliferação é uniforme, o crescimento é concêntrico.

Introduzindo um novo nível de regulação sobre os modelos lógicos, isto é, considerando que estes podem ser modulados por uma medida de densidade celular relativa às posições, o crescimento epitelial perde a uniformidade e formam-se padrões de crescimento distintos, que podem ser explorados para o estudo de morfologias ramificadas.

Finalmente, foi desenvolvido um modelo lógico para a especificação das células sensoriais precursoras dos mecanorreceptores dos organismos da ordem *Diptera*. Existem dois tipos de receptores (macro- e microrreceptores). Os padrões de especificação asseguram a não-adjacência de células precursoras através de um regulador central comum aos dois tipos de receptores. A activação deste regulador está dependente de um sinal pré-existente, que induz a formação de grupos de células competentes. Nestes grupos, a selecção de células precursoras isoladas depende de mecanismos de cooperação lateral (mediada pelo receptor EGFR), e repressão lateral (através do mecanismo de sinalização Delta-Notch), que são definidos em dois modelos lógicos (para cada tipo de receptores). Os mecanismos que levam à formação dos grupos de células competentes estão temporalmente associados a fases distintas de crescimento e desenvolvimento do organismo, pelo que este caso de estudo é adequado à validação dos métodos implementados.

As propriedades de especificação de células precursoras dentro de grupos de competencia pré-formados são facilmente recuperadas, e são semelhantes às apresentadas na literatura consultada. No entanto, na tentativa de recuperar os mecanismos que definem a distribuição do sinal pré-existente através da proliferação celular, obtiveram-se

comportamentos de dispersão celular. Estes resultados expõem as restrições do método considerado para a divisão celular e consequente reajustamento epitelial. Não tendo em conta as propriedades adesivas entre as células, os grupos de células relacionadas dispersam, e não é possível formar agregados de células competentes, a partir das quais se possa especificar uma célula precursora.

No entanto, estes resultados fundamentam as vantagens de abordagens discretas para modelar sistemas epiteliais, e definem os requerimentos mínimos à implementação de métodos neste tipo de quadros de modelação, para a inclusão de crescimento em tecidos compactos.

Acknowledgements

I want to thank my advisor Claudine Chaouiya, an outstanding investigator, and an even more extraordinary person. Her motivation, patience and constant enthusiasm made this an unforgettable experience. To Pedro Monteiro, who spent copious amounts of time helping and teaching me. To both of them I am grateful for all the support, both financial and personal.

Moreover, I thank IGC, that hosted me during this year, and all its amazing community.

To my co-advisor Francisco Dionísio, who has been an inspiration.

To Marco, for his company and long afternoon discussions on our work.

For the emotional support and company, I thank my little aunt Márcia and my grandparents.

To Jocelyn, the best of friends, my company in good and bad moments, who made me see the bright side of things when I couldn't.

And to my loving parents, who always supported me in my choices.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Due to its complexity, biology has always been an inherently observational area, where explanation of natural phenomena strongly relies on the collection of experimental data in order to extract regularities and patterns, frequently resorting to statistical analysis.

Yet, experimental data and observations do not constitute knowledge *per se*, and it is when a mechanistic process is postulated under a minimal set of underlying principles that an hypothesis is born, providing a framework to explain empirical data.

Theoretical approaches have long ago proven powerful in the assessment of physical properties through quantitative methods, but only recently, and with the aid of computational power, have they started to gain relevance in life sciences (Shou *et al.*, 2015). Unlike the often called *phenomenological* or statistical approaches, theoretical models evolve according to rules that have been abstracted from the underlying biological process, defining *e.g.* kinetic mechanisms of regulation.

Here, the importance of theoretical models comes from the fact that they can be seen as "working hypotheses" that can be unambiguously investigated, by comparisons with experimental data. From such analyses, two major outcomes can occur: if the model recovers most data properties, it may be able to provide relevant predictions, guiding investigation; otherwise, it reflects the inconsistencies

between the investigator's understanding of a system and experimental results, leading to hypothesis rejection and model refinement.

Given the current ability to generate, store and handle unprecedented amounts of data, there is an extensive need of theoretical approaches to comprehend this gigantic source of information. The choice of the modelling formalism depends on the type of information the modeller wants to extract. In particular, information sourced from molecular regulatory systems can be accurately described by quantitative methods, using non-linear differential equations. These represent the continuous evolution of rates or concentrations of the interacting molecules through time. However, for these systems two barriers occur: the frequent absence of analytical solutions (where numerical approaches must be used) and the lack of parameter estimates. This issue becomes extensive when the system under study involves a great number of regulatory components (Le Novère, 2015).

Thomas (1973) was one of the first to propose the application of the logical approach to kinetics of regulatory networks that often comprise several intertwined feedback circuits. Logical descriptions consider that each component of the network has only two levels of activity, 0 (inactive) or 1 (active).

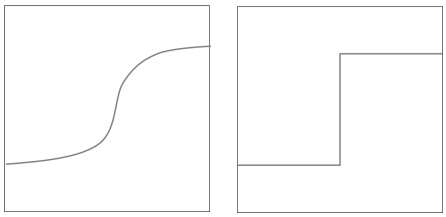


Figure 1.1: The relationship between the concentration of a substrate molecule X, and the rate of product Y synthesis, and the step-function abstracting the sigmoid relationship between the concentration of X and synthesis of Y.

The advantage of this framework is that most interactions in biology have a non-linear behaviour. Logical abstraction assumes an *infinitely non-linear idealisation* of gene regulation, which is thus considered as a discrete process depending on the presence or absence of the regulators (Thomas, 1991).

The logical framework is particularly successful to build and analyse models of cellular networks. Since the application of Boolean networks to biology, the framework has been extended and refined to account for multi-value variables, changes of the update schemes and methods have been developed to perform dynamical analysis (Abou-Jaoudé *et al.*, 2016).

Logical approaches are suited to model long term cell fate decisions that can be identified by the presence or absence of molecular markers.

Cellular automata are also discrete dynamical systems useful for the study of emergent behaviours in multi-cellular biological systems. They encompass lattices that can be representative of tissues, where each unit is a cell. Moreover, every cell is associated to a phenomenological state, usually representative of its phenotype / fate. Cell behaviours are influenced by interactions with other cells and environmental factors (such as extracellular matrix, chemicals, and forces). In their original definition, cellular automata are a subset of systems under the logical formalism, *i.e.*, the state of the cells is considered to be binary (0 or 1) and the regulation is given by a cardinality constraint over the surrounding cells (all-or-nothing rule given the neighbours state).

Cellular automata have been used to study system behaviours that can be translated into patterns, producing very sophisticated self-organised structures. One motivation for using cellular automata is the different scales on which typical biological phenomena depend. In most cases, it is infeasible to describe the behaviour of a tissue in terms of individual molecular dynamics. Cellular automata provide a natural mesoscopic scale that relates the aforementioned molecular dynamics (as the state of the cells) to the behaviour of the whole.

Pattern formation in tissues is given by complex rearrangements of cell phenotypes that depend on coordinated developmental mechanisms. These mechanisms can be interpreted as gene product interactions and changes in cellular behaviours (signalling, differentiation, mitosis, apoptosis and others) causing an arrangement of cell states that can be qualitatively distinguished from random distributions. Jernvall *et al.* (2003) consider two main developmental mechanisms: *inductive* mechanisms are defined by cell signalling mechanisms, which establish tissue patterning. *Morphogenetic* mechanisms act on a higher scale, regulating morphology of the tissues through local effects on the cells. Morphogenetic mechanisms can change the spatial distribution of cells without directly changing cell states (*e.g.*, cell motions).

EpiLog is a computational tool that integrates logical models of regulatory networks into the cells of a cellular automaton. Hence, there is an explicit representation of cell fates through the cellular networks within each cell, coupled with the definition of cell signalling associated to the automaton. The framework provides a multi-scale approach to patterning in monolayered tissues, where the

behaviours of the cells can be perspicuously understood, in an environment where the abstraction (both in signalling and network definitions) can be subject to a direct biological interpretation.

As the only framework of its kind, a systematic analysis of possible applications and prospects on potential extensions are numerous, and may help uncover developmental and disease related mechanisms.

1.2 Objectives

The core of this thesis is to develop a method that accounts for morphogenesis in the aforementioned framework. Models of tissue growth and size regulation often rely on topological models where cell shapes and sizes can be explicitly defined.

Hence, our objectives can be subdivided into:

- Perform the required corrections to the tool EpiLog (bugs, algorithms, Graphical User Interface) in order to release a stable version;
- Develop a form of signalling noise to overcome the inherent synchrony of cellular automata, and qualitatively analyse the impact of model behaviours;
- Formalise a method to account for tissue morphogenesis, while maintaining the same level of abstraction, and implement it in EpiLog;
- Develop a model on the specification of bristle mechanoreceptor cell to analyse and validate the aforementioned implementations.

1.3 Contributions

The contributions of this thesis can be dissected as follows:

- Definition of mechanisms in the logical abstraction that provide the bases for the formation of Turing patterns from pools of equipotent *naive* cells. These can further be improved and stabilised by the combined use of the previously implemented temporal asynchronism of the automata, with the

newly included signalling noise. Our results suggest that uniformity in pattern formation can be obtained through signalling noise;

- Formalisation of a compression measure modelling the shape of growing tissues through conditioned cell displacement, without resorting to topological properties of the cells;

- Assessment of basic regulatory mechanisms that lead to morphogenesis in proliferating epithelia, by affecting the spatial distribution of cells that maintain mitotic activity;

- Development of network models that account for the three main mechanisms - induction, cooperation, inhibition - involved in Sensory Mother Cell selection for the formation of bristle patterns in *Diptera* species;

- Determination of further requirements for the implemented extension suited for the modelling of pattern formation through clonal propagation.

1.4 Overview

This document is organised as follows: Chapter 2 introduces the reader to the formalism and dynamics on regulatory networks and cellular automata. Chapter 3 presents a description of EpiLog's framework, and its respective presentation in the GUI of the software. This is partially done resorting to a previously published model, developed and analysed in a first prototype of the software. Section 3.4 regards the role of asynchronism is cellular automata. The section presents the previously existing form of asynchronism. Moreover, signalling noise is formalised and implemented. The results obtained from comparison and combination of both parameters show that synchrony disruption leads to uniform stability.

The core subject of this work (Section 3.5 and Chapter 4) address respectively tissue growth and morphogenesis (and subsequent implemented methods of cell proliferation) and the case study of bristle patterning, that requires the aforementioned extensions accounting for asynchronism and proliferation. The final

chapters summarise the extent to which the framework can contribute for advances in life sciences research, as well as the challenges and required refinements to account for in the future.

Chapter 2

Formalism and Dynamics of Discrete Systems

This chapter presents the basics on the logical modelling framwork and on cellular automata.

2.1 Logical regulatory networks

A dynamical logical regulatory network, or simply Logical Model, is a mathematical object defined by a set of components associated to discrete variables representing their levels of activity or concentration. These variables are often binary (Boolean abstraction - presence or absence) but may also be multi-valued. Sets of discrete rules define the values of the variables with respect to the state of the system.

In a Logical Model the states of the components evolve step by step, without an explicit time scale, allowing for a qualitative analysis.

2.1.1 Model definition

Formally, a Logical Model is given by $(\mathcal{G}, \mathcal{K})$ where (see Abou-Jaoudé *et al.* (2016) for further details):

- $\mathcal{G} = \{g_1, g_2, \ldots, g_n\}$ is the set of n regulatory components. Each g_i is associated to an integer value max_i, defining its discrete range of functional

levels as $\{0, \ldots, max_i\}$. The full state of the model is represented by a vector g of size n giving the values of the variables. Such a state is an element of the finite set of states of the model (all possible combinations of component levels), *i.e.*, the state space of the system denoted by $\mathcal{S}$ and defined by the cartesian product $\prod_{i=1}^{n}\{0, \ldots, max_i\}$.

- $\mathcal{K}$ is the transition function defined by a set of discrete logical functions $\mathcal{K}_i$. Each logical function $\mathcal{K}_i$ univocally defines the value of g_i depending on the state g of the model. The transition function is therefore a mapping between model states ($\mathcal{K} : \mathcal{S} \rightarrow \mathcal{S}$). The transition function defines the successor state of g as $\mathcal{K}(g) = (\mathcal{K}_1(g), \mathcal{K}_2(g), \ldots, \mathcal{K}_n(g))$.

A Logical Model is associated to a regulatory structure or regulatory graph. This encompasses the set of nodes (the components in $\mathcal{G}$) and signed directed edges denoting regulatory relationships (activations or inhibitions) between the nodes. Given a Logical Model, its transition function $\mathcal{K}$ is sufficient to deduce the associated regulatory graph. The number of components defines the number of nodes in the regulatory graph, while the directed edges can be deduced from each $\mathcal{K}_i$, defining the regulators of the component i. However, there may exists several sets of transition functions compliant with a regulatory structure. Hence a regulatory graph alone may define a family of Logical Models (and not necessarily a unique model).

Nodes with no incoming edges possess no logical functions. They are the input nodes of the system (e.g. membrane receptors), whose regulation is considered to be external. Their values are thus fixed.

2.1.2 Model dynamics

A Logical Model defines a discrete dynamics over the state space $\mathcal{S}$. The dynamics is defined by the transition function $\mathcal{K}$. Given a state g, if $\mathcal{K}(g) \neq g$, there is at least one g_i whose value is not compliant with the logical rule (*i.e.* it has an update call, $\mathcal{K}_i(g) \neq g$), and the system state changes. Otherwise, if $\mathcal{K}(g) = g$ then the state is stable, and there are no update calls for any g_i.

The dynamics of these models are conveniently represented by State Transition Graphs (STG), where each node represents a state of the system, and each directed edge a transition between two states. Because $\mathcal{S}$ is finite, models' dynamics always end up in a terminal Strongly Connected Component (SCC). The simplest form of a SCC is a point attractor or stable state (Subsection 2.3 contains further information on the subject).

Several updating schemes have been considered in the literature (see Abou-Jaoudé *et al.* (2016)). These define the successor states, or more precisely, how updates of component variables are combined. The most common schemes are the synchronous and the asynchronous updates, described below.

In *synchronous* dynamics, the transitions between states define the simultaneous changes of values for all the components called for update. Here, the dynamics are deterministic, each state has at most one successor, and any initial state defines a single trajectory.

In *asynchronous* update schemes, each component is updated independently, yielding non-deterministic dynamics. When two or more variables are called to update, these changes are said to be concurrent. Asynchronous dynamics consider the calls as independent updates of the level of the variables. In this context, the *Random asynchronous* update scheme defines a probability (often uniform) to choose one of the possible transitions among the concurrent ones. At each step, only one component variable is updated (Chaves *et al.*, 2005). Furthermore, one can consider the *Fully asynchronous* dynamics in which all possible combinations of concurrent transitions and their associated probabilities, selecting one of the combinations (Bahi & Contassot-Vivier, 2002).

In the (full) asynchronous schemes, the STG contains concurrent trajectories, although only one of them may be actually chosen by the system. Note that different updating schemes possibly lead to different attractors (only stable states are conserved) and reachability properties. Asynchronous dynamics are more realistic than the synchronous ones, which often generate spurious cycles.

The SCC are the model attractors that define long term behaviours, often associated to differentiated cellular states. In addition to their identification, it is useful to study the reachability properties of these attractors, *i.e.* from an initial

state, what is (are) the reachable attractor(s), what are the properties along these trajectories, etc (Abou-Jaoudé *et al.*, 2015).

Transitions between states of the system can also be changed through the definition of priority classes. These denote a qualitative difference in the time scales at which updates to the components values can occur (*e.g.* the relative speed of translational and post-transcriptional processes, or the difference in the velocity of activation and repression of a component) (Abou-Jaoudé *et al.*, 2016).

Formally, for a set of regulatory components $\mathcal{G}$ of size n, it is possible to define p priority classes $C_1, C_2, \ldots, C_p$. To each of these classes a rank and an updating scheme (synchronous or asynchronous) is attributed. The rank defines a relative update speed, *i.e.*, given a set of components distributed among classes with different ranks, at each time step, the components in the highest ranked class are checked for updates. If they are updated, the components in the lowest ranked classes remain unchanged. These will only be assessed for updates when there are no updates to the components in the higher ranked classes. These definitions account for knowledge on kinetics, preventing unrealistic trajectories to occur.

2.1.3 Model perturbations

In Logical Models two main forms of perturbations can be considered to disrupt the canonical behaviour of the system. The first form of perturbation acts on the model components by forcing their corresponding variables to take specific values regardless of the logical functions. This model perturbation can be associated to gene knock-outs or ectopic expressions. The second one is subtler, and considers an alteration of the logical functions as a mean to simulate loss or gain of sensitivity to regulators or binding activity to specific substrates. Both forms of model perturbations modify the function $\mathcal{K}$.

Furthermore, perturbations provide a systematic approach to formulate novel predictions and guide investigations.

2.2 Cellular automata

Cellular automata (CA) are a class of discrete dynamical systems to model multi-agent systems in which emergent behaviours result from local interactions.

A cellular automata encompasses an n-dimensional array of elements - cells - representing the individuals, to which a finite set of states can be associated. The number of considered states is usually small (no more than three or four). Macroscopic behaviours emerge from the effects of local interactions amongst neighbouring elements. The state of an individual element changes with respect to rules that depend on the states of neighbours and of its own state.

2.2.1 Model definition

Formally, a cellular automaton is a mathematical object defined by $\{\mathcal{L}, \mathcal{S}, \mathcal{N}, f\}$, where (Bouré *et al.*, 2011):

- $\mathcal{L}$ is the n-dimensional array of the cellular space (lattice) of fixed size. Without loss of generality, we consider $\mathcal{L}^{m \times n}$ a $2D$ automaton, where each element c of the lattice is referenced by its coordinates as c_{ij} ($i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\}$);

- $\mathcal{S}$ is the (non-empty) finite set of states each cell can adopt. We consider that retrieving a cell c_{ij} in the lattice is the same as retrieving its state;

- $\mathcal{N}$ is a collection of relative positions (a, b) (with $a, b \in \mathbb{Z}$) that associates each cell to a number of interacting neighbours. Given a neighbourhood $\mathcal{N} = \{n_1, n_2, \ldots, n_k\}$, the neighbours of a cell c are the set $\{c + n_1, c + n_2, \ldots, c + n_k\}$;

- f is the local transition (or update) rule that defines the new state of each cell accordingly to its state and to the states of its neighbours.

A configuration of the system is given by a function x that maps every cell i, j in the lattice $\mathcal{L}$ to a state of the state space $\mathcal{S}$.

Logical Models and Cellular Automata represent particular cases of the same generic network structure. But they present distinct specificities: Cellular Automata are well suited to model large systems efficiently and in a robust manner, even though complexity (possible configurations) increases exponentially with the size of the cell space. CA represent expandable networks (different lattice sizes) whose components are identical, usually dependent on a high number of regulators (the neighbours). Dynamical analyses of such systems and attractor detection become difficult when compared to Logical Models, which although diverse in components and interactions, are frequently smaller and allow for more extensive analysis.

2.2.2 Model dynamics

The evolution of a Cellular Automaton relies on the global transition function computed through the local transition functions. For a given configuration x^t of the system at step t, the global transition function $F(x^t)$ determines the configuration at $t+1$ (*i.e.* x^{t+1}).

Therefore, considering $\mathcal{N} = \{n_1, n_2, \ldots, n_k\}$, the function $F(x^t)$ is defined by:

$$\forall c \in \mathcal{L}, x^{t+1}(c) = f(x^t(c), x^t(c+n_1), x^t(c+n_2), \ldots, x^t(c+n_k)).$$

For every cell in the automata, its next state is computed synchronously: all cells are updated at the same time.

As in the Logical Models, terminal SCC always exist. If $x^t = x^{t+1}$, the CA has reached a stable pattern, analogous to the Logical Model stable state. A pattern is stable if no cells have updating calls.

The number of stable configurations may be incredibly large (making it impossible to compute all of them by brute force), but many of them can be similar by applying symmetry or rotation operations.

In simple automata (binary nearest-neighbour 1D automata), Stephen Wolfram defined four classes of behaviour and systematically classified the 256 possible rule combinations in what became know as the Elementary Cellular Automata (ECA). He proposed four main classifications of the different types of dynamics,

distinguishing forms of chaotic, stabilising and cyclical dynamics (Gramß *et al.*, 2005).

For more complex CA models it is not possible nor meaningful to transpose this classification. For a small $\mathcal{S}$, temporal analysis of the evolution in the number of cells at each state provides insightful information about the system. For instance, in *SIR* models (Susceptible-Infected-Recovered epidemiology models) a three state automata is considered. A cell in the S transits to I as a function of the number of neighbours in I. This is performed in accordance with a probability distribution that reflects the virulence of the infection. Afterward it moves to R and can remain at that state for a specified amount of steps before returning to S (partial immunity). Stability detection is not so important when compared with temporal analysis of the evolution of the number of cells in the I state, which accurately models endemic and epidemic behaviours of infectious diseases. For endemic diseases, the I cells never disappear but their numbers suffer regular oscillations. Hence, a stable state does not necessarily arise but periodic oscillatory dynamics are detected (White *et al.*, 2007).

A measure of the energy of stable pattern in CA energy is often used. The energy of a stable pattern is defined by the number of contacting cells in the same state. The higher the number of contacting neighbours in the same state, the more uniform the stable pattern is, and the energy of the system lowers (Regnault *et al.*, 2008).

2.2.3 Model perturbations

Perturbations is CA can be made by altering local transition functions. Kim *et al.* (2009) demonstrated this application in a discrete automaton to model cystic organoid formation. Their model consisted in a grid of hexagonal cells. Each element could be in one of three states, stating that it would either be a space occupied by extracellular matrix, luminal space, of a biological cell. Twelve local transition rules with biological basis were defined, and the model was able the recover normal cystogenesis development. To compare their results with experimental perturbations, the authors systematically dysregulated the outcomes of the updating rules, and mapped the mutated growth and morphology to known

altered phenotypes. This approach allowed for the prediction of which signalling mechanisms contributed the most to pathological conformations related to cancerous tissue formation.

Other forms of perturbations include the specification of fixed points in the array, *i.e.* sets of cells that disregard the updating rules of the automata, working as point mutations that affect the state of "wild-type" neighbours and resulting stable states. Changes to the sensitivity to signals (without changes to the outcome) were also proposed as means of simulating loss of function or gain of function mutations (de Back *et al.*, 2012).

2.3 Remarks on attractors and their reachability properties

The dynamics of any finite discrete dynamical system is eventually trapped in a terminal maximal set of mutually reachable states (a terminal Strongly Connected Component), from which there are no outgoing transitions. The simplest form of a terminal SCC is a point attractor or stable state. Complex attractors containing two or more mutually reachable states can subdivided into two types: elementary cycles, which are closed transitions with exactly one outgoing transition for each state, and non-elementary cycles where each state can have more than one predecessor or successor (*i.e.* a combination of cycles).

Properties of interest reside in these attractors, and in how they can represent cellular long term behaviours. However, besides the detection of an attractor another issue concerns its reachability, *i.e.* what are the states of the system for which there is a path leading to that attractor. The existence of point attractors is independent of the update scheme. Nonetheless, the choice of the updating scheme affects the reachability of a point attractor. The nature of complex attractors depends on the update scheme. As it has been previously mentioned in Subsection 2.1.2, synchronous update schemes generally lead to cycles with no biological meaning. Different update schemes generate different possible transitions. In particular, transitions in a complex attractor may only exist for specific update schemes.

The *basin of attraction* of a terminal SCC is the set of states whose trajectories lead univocally to that SCC. In the synchronous update scheme the trajectories are deterministic, *i.e.*, there is at most one outgoing transition from one state to another until the SCC is reached (Figure 2.1 (a) and (b)). In asynchronous update schemes this may not be true, as there may exist several outgoing transitions for a state (Figure 2.1 (c)).

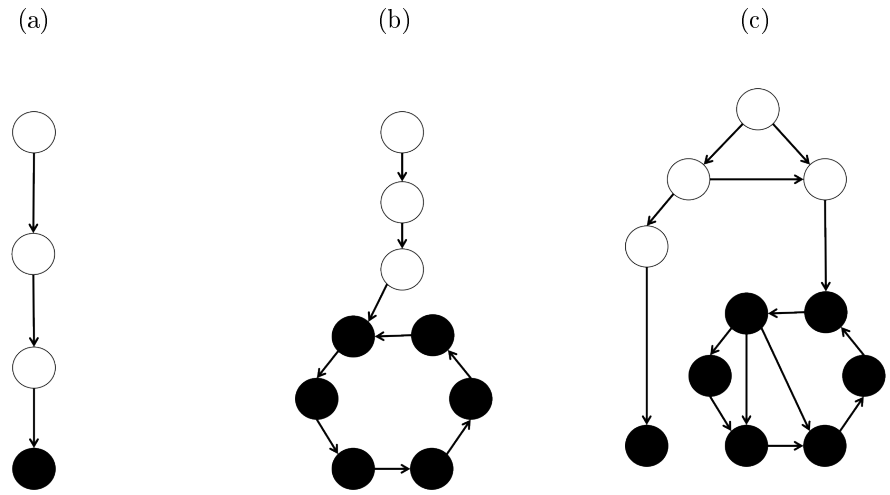


Figure 2.1: Possible evolutions of discrete dynamical systems from an initial state. Nodes represent states of the system, and directed arrows are the transitions between states. Black nodes are the states that compose the attractors. (a) Dynamics of a synchronously updated Logical Model, converging to a point attractor. (b) Dynamics of a a synchronously updated Logical Model converging to an elementary cycle. (c) Asynchronous state transition graph in which each state of the system can have more than one outgoing transition. Here, several terminal Strongly Connected Components can be reached (a point attractor and a complex attractor).

Unreachable stable states are singularities of a larger group of states for which there is no predecessor state, the *Gardens of Eden*, and therefore can only occur if they are the initial state. Hence, because update schemes affect transitions, Gardens of Eden also depend on the update schemes.

Furthermore, it may be relevant to study the conditions under which the modelled system can leave an attractor. By applying perturbations or by changing the values of the input nodes in a stable configuration, it is expected that a so-called reprogramming event (in differentiation processes) may occur. This type of analysis can provide insights into strategies to modulate cell behaviours, *e.g.* in pathological conditions. These predictions are guiders for experimental work.

Chapter 3

EpiLog

3.1 General description

EpiLog[1] is a freely available software tool for the development of multi-level **Epi**thilium **Log**ical models. The tool implements a Cellular Automata (CA) framework in a 2D-grid of hexagonal cells. To each cell, a Logical Model can be associated, directing the cell behaviour. From hereon, the use of the term cell will be equivalent to a reference to its Logical Model and respective state.

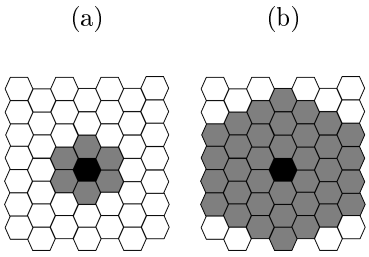


Figure 3.1: Neighbourhood examples relative to the black cell. Neighbourhood **(a)** only considers contacting neighbours. Neighbourhood **(b)** is specified by the range $[1, 3]$.

Here, the CA cells state space is the state space of the Logical Models associated to them. In addition, it is possible to define more than one neighbourhood. Each different neighbourhood respects a different range of distances measured in cell units. An individual neighbourhood is defined by a range of distances defining a radius of interaction, as depicted in Figure 3.1. Therefore, it is represented by an inclusive range $[d_{min}, d_{max}]$ where d_{min} is usually 1.

In the absence of communication between the cells, each Logical Model converges to an attractor as the result of its initial state and transition rules governing

[1] http://epilog-tool.org

its evolution synchronously. Cell communication is defined as an effect of neighbouring cell states on the input nodes of the Logical Models associated to the cells. As mentioned in Section 2.1, input nodes in Logical Models depict components for which the regulation is external to the cellular network. Therefore, in this framework, the level of such components can be constant (a spatially distributed microenvironmental cue that remains static throughout simulations) or dynamic, as a function of signals sent by neighbours.

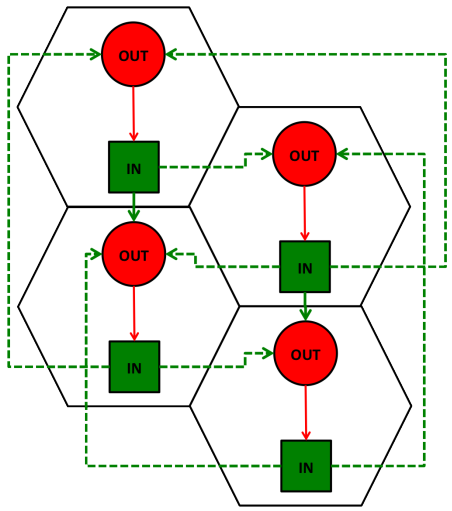


Figure 3.2: Example of integration of a Logical Model in a CA. The intra-cellular Boolean model contains one input node *Out*, and one internal node *In*, which is inhibited in the presence of *Out*. Communication (dashed lines) defines that the *Out* nodes will be active if at least one of the direct neighbouring cells has the *In* node active:
$Out = 1 \;\; if \;\; In(1\ ,\ 1 : \_\ ,\ 1)$

It must be ensured that the regulators (component instances in neighbouring cells) are not input components of the Logical Models. This requirement is interpreted under the biological scope as the effect of a signalling molecule produced by a cell (an internal component of the Logical Model) on a receptor.

Cell communication occurring within tissues to form complex patterns mostly relates to juxtacrine (membrane bound) or paracrine signalling (mediated by short range diffusion or cytoplasmic extensions) (Přibyl *et al.*, 2016). If signals are diffusible, different ranges of effects reflect the diffusion coefficient of those molecules. Hence, the framework considers the extension of the neighbourhood to account for this fact. If the neighbourhood is defined by $d_{max} = 1$, the signalling is contact dependent (juxtacrine). Otherwise, if $d_{max} > 1$, the signal is usually a diffusible molecule.

Formally, the transition rules (here called *integration functions*) acting on Logical Model inputs are conjunctions, disjunctions and negations of atoms depicting cardinality constraints over one signalling molecule. A cardinality constraint is represented in the form:

$$x_I = S(t_{min} : t_{max} \ , \ n_{min} : n_{max} \ , \ d_{min} : d_{max})$$

where:

- x_I is the level of the Logical Model input I;
- S is the Logical Model component acting as a regulatory signal;
- $t_{min} : t_{max}$ is the range of levels of the signal affecting the value of x_I. If t_{max} is left unspecified, only t_{min} will be considered. If t_{max} is replaced by "_", all possible levels of S above t_{min} are considered;
- $n_{min} : n_{max}$ is the constraint over the number of neighbouring cells for which the level of component S is in the specified range. As before, if n_{max} is replaced by "_", the upper bound of this constraint is the total number of neighbours;
- $d_{min} : d_{max}$ is the radius range of the neighbourhood. If d_{max} is not specified, only cells in the radius defined by d_{min} will be considered a neighbourhood.

Figure 3.2 provides an example of communication between cells and associated Logical Models. This "toy model" will be thoroughly explored in the next sections. It is a simple 2-state Logical Model, in which an input component *Out* negatively affects the expression of the internal component *In*. In the Logical Model, there is no regulation of the *Out* component, but in an epithelial context its regulation is provided by signals from surrounding cells. If the *Out* component is highly sensitive to juxtacrine signalling, it is only required that at least one direct neighbour contains *In* at level 1. The integration functions represent a qualitative interpretation of the sensitivity to neighbouring signals. For any form of interaction (positive or negative) between signals and input, the sensitivity of the input to the signal is inversely proportional to the required minimum number of cells emitting that signal.

Simulation is performed in a stepwise manner by simultaneously updating the state of the cells with respect to their previous state and to the previous state of their neighbours (and by extent, of their signals), until a terminal SCC is

reached. This can either be a stable pattern (no cell Logical Model has further state transitions) or a terminal oscillation. Detection of terminal oscillations is hard because in the most extreme case, all the states of the automata are part of it. For any array $\mathcal{L}$ of size $m \times n$, the number of possible states is $\mathcal{S}^{m \times n}$, increasing exponentially with the size of the grid.

Each step of the simulation can be visualised on the grid and the state of the system can be assessed through a system of colours. The default colour of the cells is white. Every component of the Logical Model has an associated colour. If the component is multi-valued (*i.e.* several levels of activity), the highest level of activity corresponds to the specified colour, while lower levels have lighter shades. When one component of the Logical Model is selected, all cells expressing that component change from white to the component colour. If there is more than one selected component, the cell colours are the combination of the component colours.

This approach allows for an easy comparison between computational simulations and experimental data, *e.g.* selecting the same components for visualisation as the corresponding fluorescent targets used in microscopy.

3.2 Extensions

EpiLog couples the CA framework with the Logical Model framework. Therefore, some of the presented methods for Logical Models can be transposed for this tool, and in particular, to individual cells of the automaton.

Distinct initial conditions can be applied to each cell, *i.e.* for each cell it is possible to define the initial level of the Logical Model components. This can be set for all the variables except the input nodes whose values are regulated by cell communication.

For each Logical Model in use, it is possible to define the updating orders as proposed in Subsection 2.1.2. The established relative priorities between component updates are applied to all the cells in the grid.

Furthermore, modellers can consider Logical Model perturbations (Subsection 2.1.3) in individual cells of the automaton. These cells act as mutated elements of the population, and affect the resulting behaviour of its neighbours, as well as

the subsequent stable patterns. This approach is particularly interesting for the detection of aberrant phenotypes and relationship with signalling mechanisms.

The grids can be subject to boundary continuity definitions, by stating that cells in opposite corners are adjacent to each others. Hence, vertical or horizontal boundary conditions define a hollow cylinder. Combined vertical and horizontal continuity generates a toroidal grid.

3.3 Application: Dorsal appendages model

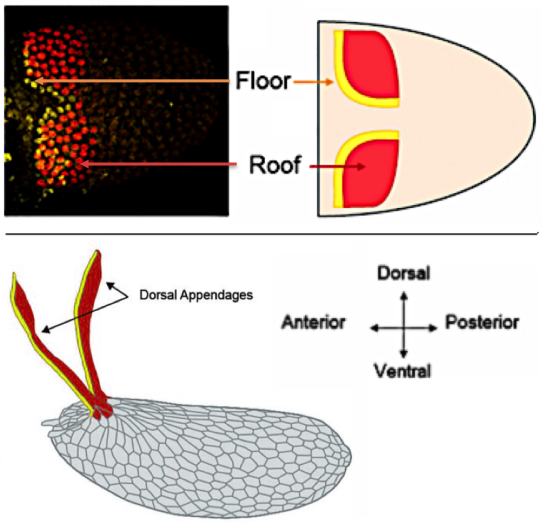


Figure 3.3: Dorsal appendages patterning and morphology. The specification of floor and roof cells (upper panel) later results in prominent dorsal appendages in the dorsal anterior of the egg. Figure adapted from Zartman *et al.* (2011).

EpiLog has already been proven to be useful in assessing pattern formation that results from complex regulatory circuits in epithelial contexts. Fauré *et al.* (2014) used a first prototype of the software to unveil the mechanisms of Dorsal Appendages (DA) cell specification in the *Drosophila* eggshell. The structures are thought to facilitate gas exchanges of the burrowed eggs.They arise from an already established pattern. The authors recapitulated the patterning of the eggshell through the involved cellular networks and cell-cell communication modulating the Logical Models.

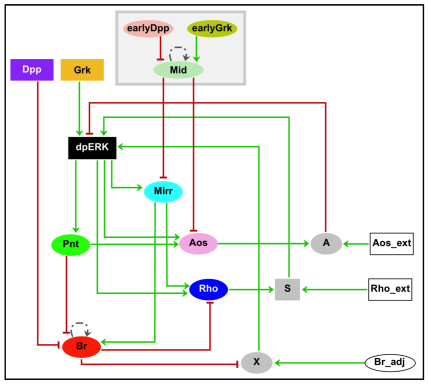
The pattern consists in two spots of *broad* (*br*) expression along the dorsal midline (present during oogenesis late stages) and respective adjacent expression of *rhomboid* (*rho*). The expression of *br* determines the DA roof, while *rho* is associated to DA floor cells. The network model was developed in GINSim[1], and considered two main signalling pathways: EGF pathway, where the EGF receptor is activated through the TGF-like ligand Gurken (Grk), and the BMP pathway, regulated through the gene *decapentaplegic* (*dpp*). Furthermore, the floor or roof cell decision is spatially heterogeneous, and only cells in the anterior region are competent to commit to that fate. Hence, the authors considered an additional regulator of the cellular network by accounting for an anterior-posterior compartmentalisation signal, the transcription factor Midline (Mid).

[1] http://ginsim.org/

The regulatory model was imported to EpiLog, and the eggshell cells were represented in a 20 by 40 grid. Here, the initial conditions were defined by heterogenous expression of Grk, Dpp (*decapentaplegic* gene product), and Mid over the otherwise *naive* cells (*i.e.* the remaining components have their value set to zero). The implemented framework allows for this through a graphical interface, where components of the Logical Model and respective levels of activity can be selected and applied over the cells in the grid by "painting" them. Extensive communication was defined through regulatory signalling over the Aos_ext, Rho_ext, and Br_adj input nodes, respectively through Aos, Rho, and Br in surrounding cells. Aos_ext integrates long-range information. Rho_ext and Br_adj only account for contacting cells information. Moreover, priority classes separated fast and slow updates to the components (see Fauré *et al.* (2014) for details).

The model was able to recover the eggshell patterning in wild-type condition. Moreover, simulation of clonal perturbations (also defined by "painting" the cells) were able to reproduce known mutations (depicted in Figure 3.4, bottom right panel). The model exposed inconsistencies in the literature and provided guidelines to direct experiments to unveil further details on the eggshell patterning.

(a)



Component	Level	Logical function
Mid	1	Mid | (earlyGrk & !earlyDpp)
A	1	Aos_ext:2 | (Aos & Aos_ext:1)
S	2	Rho:2 | Rho_ext:2
	1	(Rho:1 | Rho_ext:1) & !(Rho:2 | Rho_ext:2)
X	1	!Br & Br_adj
Aos	1	Pnt & dpERK & !Mid
Br	1	(Mirr | Br) & !Dpp & !Pnt
dpERK	2	[Grk:3] | [Grk:2 & ((S:2 & !A) | X)] | [Grk:1 & ((S:2 & !A) | (S & X))] | [!Grk & S:2 & (!A | X)]
	1	[Grk:2 & ((S:2 & A & !X) | (!S:2 & !X))] | [Grk:1 & ((S:2 & A & !X) | (S:1 & !X) | (!S & X))] | [!Grk & S:1 & (!A | X)]
Mirr	1	dpERK & !Mid
Pnt	1	dpERK:2
Rho	2	dpERK:2 & Mirr & !Br
	1	dpERK:1 & Mirr & !Br

(b)

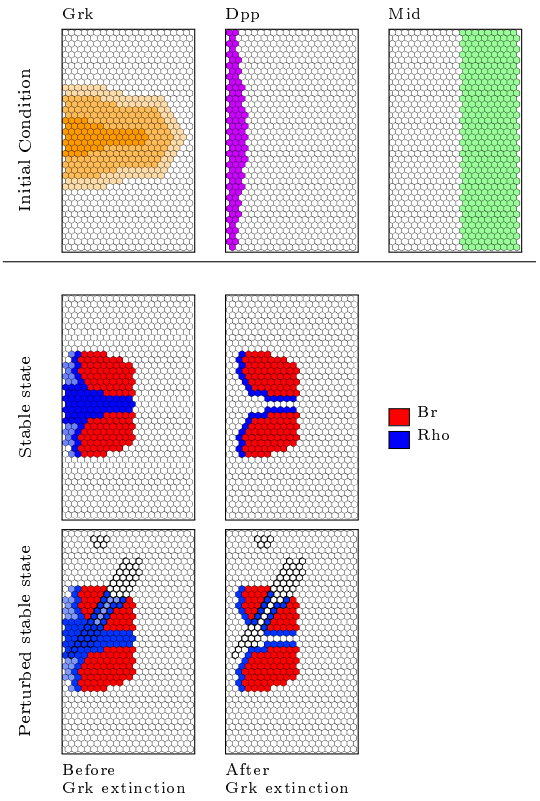


Figure 3.4: Dorsal appendages regulatory graph and embedding in the CA. (a) DA Logical Model. Oval nodes are Boolean, rectangular nodes are multi-valued. Grey nodes combine para- and/or autocrine signals. The grey box contains the module that defined the anterior competence region. Dotted edges account for protein maintenance. EGF activity is represented by dpERK. (b) Dorsal appendages integration in CA and stable states. Initial state of the cells is naive, upon which Grk, Dpp and Mid signals are applied, as depicted in the upper panel. The colour scheme is the one used in the Logical Model. Different shades of Grk depict different levels of activity. After reaching a stable state, the Grk signal is set to zero (experimental observations suggest its extinction), and the simulation proceeds. The wild type patterning of the eggshell is recovered (middle panel). The application of cell specific perturbations, denoted by bold borders of the hexagons, illustrate how mutations (in this case, Pnt ectopic expression at level 1 in the bold cells) affect the resulting patterns, causing altered phenotypes.

3.4 The role of stochasticity in CA behaviour

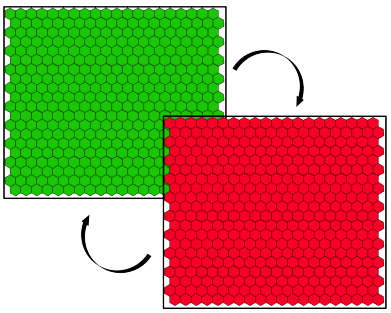


Figure 3.5: Steps of the simulation of the system described in 3.2, starting from a state where all cells have both components at zero. A terminal cycle is obtained and no stable pattern arises. The oscillation is maintained for any changes to the sensitivity of *Out* and neighbouring radius. A 20×20 grid with toroidal conditions was used.

A common theme in developmental biology is the formation of complex patterns from a field of equivalent precursor cells. Patterning structures can result from an extensive communication between cells, where non-autonomous negative feedback mechanisms play a central role (*i.e*, lateral inhibition). The sensitivity of these down-regulating signalling cascades is responsible for the fine tuning of the patterns and distribution of the cell types.

Hence, patterns under these conditions can either occur through cell cooperation or competition. In the last case, stochastic variations of signalling perception result in formation of distinct complementary subpopulations. Turing hypothesised that patterns such as pigmentation in animals, branching in trees and skeletal structures, are reflections of inhomogeneities in underlying biochemical signalling (Maini *et al.* (2012), Turing (1952)).

Cellular automata do not consider heterogeneities in signalling between cells in their definition, which constitutes a major drawbacks of their use for biological purposes. Figure 3.5 illustrates this observation for the two-state system presented in Figure 3.2. The cells are red when their *On* component is active, and green when their *In* component is active. In the synchronous case, no stable pattern arises, and a spurious cycle occurs. Therefore several documented constructs have been added over the existing definitions to tackle this problem.

Two processes define the contribution of a cell to the CA dynamics: the perception of its neighbours and subsequent update of its state, followed by the transmission of its new configuration to the neighbours also scanning their surroundings.

Disruptions to the synchronism in either of these processes can guide changes to the update scheme in CA. Therefore two classes of disruption mechanisms can

be defined (Bouré *et al.*, 2011):

- Update disruption: the update of the CA can be restricted to subpopulations of cells at each step. The most elaborate way of doing so is by attributing an internal clock to each individual cell, in the form a statistical distribution that defines the update probability of that cell. This probability increases with each step the cell is not updated. The simplified approach considers a uniform probability of a cell update at each iteration of the system.

- Signalling noise: this mechanism implements a failure to communicate the cell true state to updating neighbours. As such, the signal perception by a cell may differ from the true signal, causing it to evolve into a state that otherwise would not occur. Signal noise has been implied in “optimisation" of sense patterns. This means that qualitatively the patterns are the same, but the conformation of each cell with respect to its neighbours is made in such way that the system is more robust to occasional signalling failures (Bouré *et al.* (2011), Cohen *et al.* (2010)).

This framework implements for two forms of stochasticity to account for the aforementioned observations.

3.4.1 α-asynchronism

α-asynchronism presents a disruption on the simultaneous update of all the cells in the grid, based on the implementation proposed by Bouré *et al.* (2011). Here, α is the *synchrony rate*, and takes any values in the range $[0, 1]$.

The value of α defines the fixed proportion of cells in the grid that are called to update simultaneously at each step of the simulation (the remaining cells stay in their current state). The cells are randomly selected, and therefore may already be at a stable state. Hence, the number of updatable cells can be smaller than the number of cells called to update.

If $\alpha = 0$ the system is randomly asynchronous, and a single cell will be called to update at each time step.

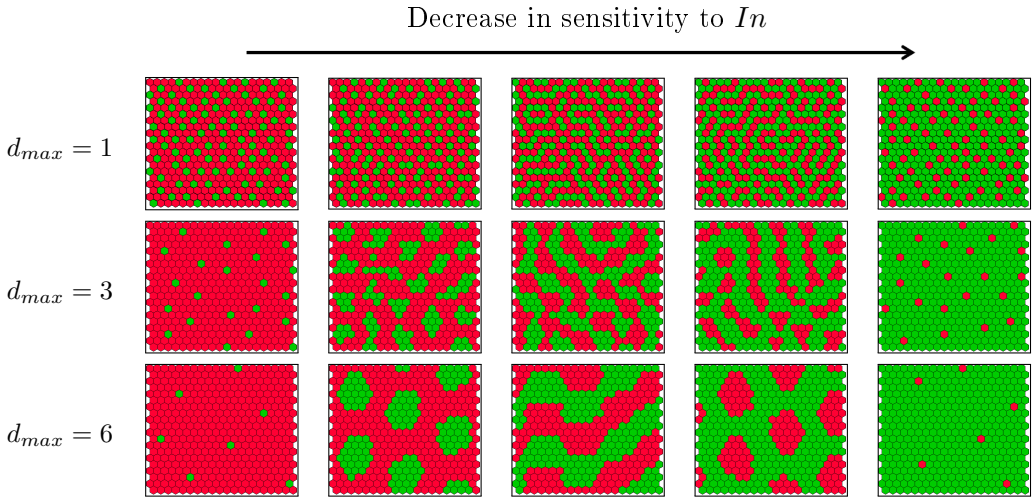


Figure 3.6: Resulting stable patterns from asynchronous update scheme applied over the system described in Figure 3.2. Each row corresponds to a different neighbourhood $[1, d_{max}]$. In each column the sensitivity of the *Out* component to neighbouring *In* signals was decreased. The patterns where obtained with $\alpha = 0.1$ (at each iteration, 10% of the grid cells are called to update).

The differences between the synchronous and asynchronous simulations (Figure 3.5 *vs* Figure 3.6) show that distinct update schemes change the reachable stable patterns and resolve spurious cycles. For either of the update schemes, the stable patterns obtained in Figure 3.6 are valid with respect to the communication function. Nonetheless, with the synchronous simulation they cannot occur from an equipotent pool of *naive* cells.

Furthermore, our results are in agreement with the observation that Turing patterns can occur from hypothetical reactions involving an activator and a repressor, favouring order in non-linear dynamical systems (Maini *et al.* (2012), Cohen *et al.* (2010)). The types of Turing patterns obtained from an activation-inhibition mechanism result from differences in the sensitivity to the regulatory signal. Five basic patterns can be withdrawn from this.

For any distance d_{max}, defining the neighbourhood range $[1, d_{max}]$, the total number of neighbours n_{total} it contains is given by

$$n_{total} = \sum_{d=1}^{d_{max}} 6 \times d = 3 \times (d_{max}^2 + d_{max})$$

Given $n_{min} \in \{1, n_{total}\}$ the minimum number of neighbours that must express In in order to activate Out in a cell, the possible stable patterns of the CA can be qualitatively classified with respect to n_{min} as follows:

$$\begin{cases} n_{min} \ll \frac{n_{total}}{2} & \text{isolated clusters of } In \text{ expressing cells} \\ n_{min} < \frac{n_{total}}{2} & \text{hybrid state: clusters and stripes of } In \text{ expression} \\ n_{min} \approx \frac{n_{total}}{2} & \text{complementary stripes of } In \text{ and } Out \text{ expression} \\ n_{min} > \frac{n_{total}}{2} & \text{hybrid state: clusters and stripes of } Out \text{ expression} \\ n_{min} \gg \frac{n_{total}}{2} & \text{isolated clusters of } Out \text{ expressing cells} \end{cases}$$

In the present case, where the initial state of the cells is the *naive*, the proportion of In and Out expressing cells in the stable pattern is bounded by the integration function. However, due to the stochastic update of the cells under the α-asynchronism, the resulting patterns may differ locally, despite maintaining its qualitative overall features (clusters *vs* stripes).

3.4.2 σ-asynchronism

σ-asynchronism was developed under the observation that signals emitted by cells can be stochastically delayed. It has been mentioned previously that cell communication is defined by a cardinality constraint over signalling molecules. These components, the integration regulators, upon changes to their states, can fail to immediately demonstrate it. Signalling failure is individually defined for each regulatory signal S. As in α-asynchronism, a variable σ_S (in the range $[0, 1]$) is associated to the signal S. At each step of the simulation a set of cells is randomly selected, whose size is given by the proportion σ_S of the lattice size. The cells that are part of this set will unveil the current value of their component S, while cells outside this set will present the value of the variable they had in the previous simulation step.

Other authors (Cohen *et al.* (2010), Bouré *et al.* (2011)) consider stochastic effects as a signal transmission failure (setting the signal to zero) instead of a delay. We discarded this approach because stable states cannot be sustained. Under a

constant disruption of signal emission, cell states are no longer compliant with the cardinality constraints, and change states.

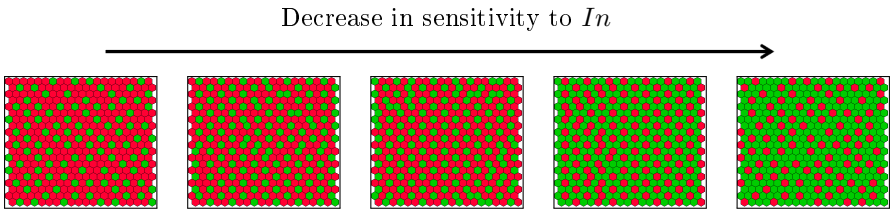


Figure 3.7: Stable patterns obtained for contacting neighbourhoods ($d_{max} = 1$), with a decrease in Out sensitivity to In, and $\sigma_{In} = 0.5$ (at each iteration 50% of the cells fail to communicate their current state). The pattens are very similar between themselves, and dissimilar from the ones obtained with α-asynchronism. No results are shown with larger distances because σ-asynchronism alone cannot lead the system to stable patterns for larger neighbourhoods.

Our prediction is that the coupling between both types of stochasticity qualitatively changes the stable patterns by changing the reachability probabilities of each pattern.

3.4.3 Results

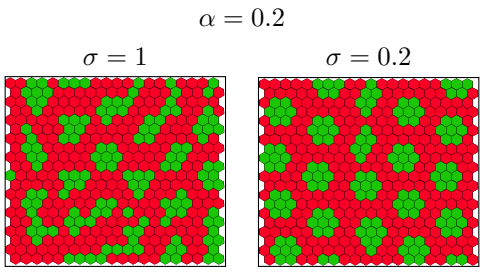


Figure 3.8: Example of the effects of signal delays in stable patterns for the same conditions. Out is activated if at least 12 of the 36 neighbours in the range $[1, 3]$ have In activated. All the cells are initially *naive*.

Signalling stochasticity cannot recover the stable patterns for complex interactions (large distances and high thresholds). Nonetheless, when coupling it with α-asynchronism, σ delays can affect the reachability properties of the system, favouring some stable patterns in detriment of others. More specifically, signalling noise is related to the smoothening of patterns. For a given sensitivity value, the distribution of Out cells relative to each other has an empirical optimal distribution that is reflected in the cluster shapes and aligned stripes. Cohen *et al.* (2010) and Bouré *et al.* (2011) have devoted some work to analyse similar properties of self-organising systems.

To assess the effects of stochastic delays, we performed a quantitative comparison of stable patterns under α-asynchronism, against its coupling with σ-asynchronism. We simulated a 50×50 grid with toroidal boundary conditions and focused on the stable patterns. To measure the properties of the stable patterns, we stored the distribution of *Out* cells with respect to the number of neighbours they have in that same state. This gives us a distribution of cell counts, separated into seven classes, *i.e*, the number of same-state neighbours as that of a cell expressing *Out* can vary between 0 (isolated *Out* cell) and 6 (*Out* cell is surrounded by *Out* neighbours).

Fixing $\alpha = 0.2$ and the neighbouring range 1 to 3, the sensitivity parameter (n_{min}) was changed from 1 to the full size of the neighbourhood ($n_{max} = 36$). For each level of sensitivity we preformed 100 stochastic simulations and generated the previously mentioned collections of data. This process was afterwards repeated with addition of the σ-asynchronism ($\sigma_{In} = 0.2$).

The distribution of the number of neighbours in the same state depends on the sensitivity to In. If the sensitivity to In is too low, *e.g* it is required that all neighbours express In in order to activate *Out* in a cell, no two *Out* cells arise together. Hence, there is an intuitive notion of "permissiveness" to the number of neighbours in the same state, in stable conditions.

Comparisons between stable patterns with and without signalling delays revealed that the number of cells having low numbers of neighbours in the same state as themselves is affected by the σ-asynchronism. This mechanism maximises the number of adjacent cells in the same state, within the ranges allowed by the sensitivity parameter. Therefore, in Figure 3.10, when In threshold is increased, the number of adjacent *Out* cells decreases. σ-asynchronism imposes an effort to maintain the adjacency.

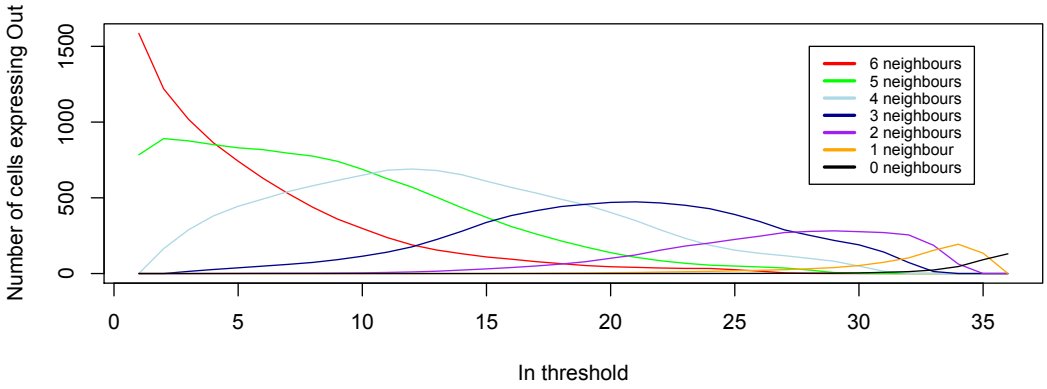


Figure 3.9: Average distribution of *Out* expressing cells containing n neighbours as a function of In thresholds

α-asynchronism patterns are more irregular because the random selection of cells to update leads to different fixed points that no longer readjust to the evolution of the automata. σ-asynchronism provides an additional restriction dependent on the fact that the higher the number of neighbours respecting the cardinality constraint in communication, the more robust it is signalling delays. Otherwise, if the number of cells respecting the sensitivity parameter is close to the turn point that changes the value of the input node, the robustness is lost and it becomes highly probable that the cells will change their state and leave stability.

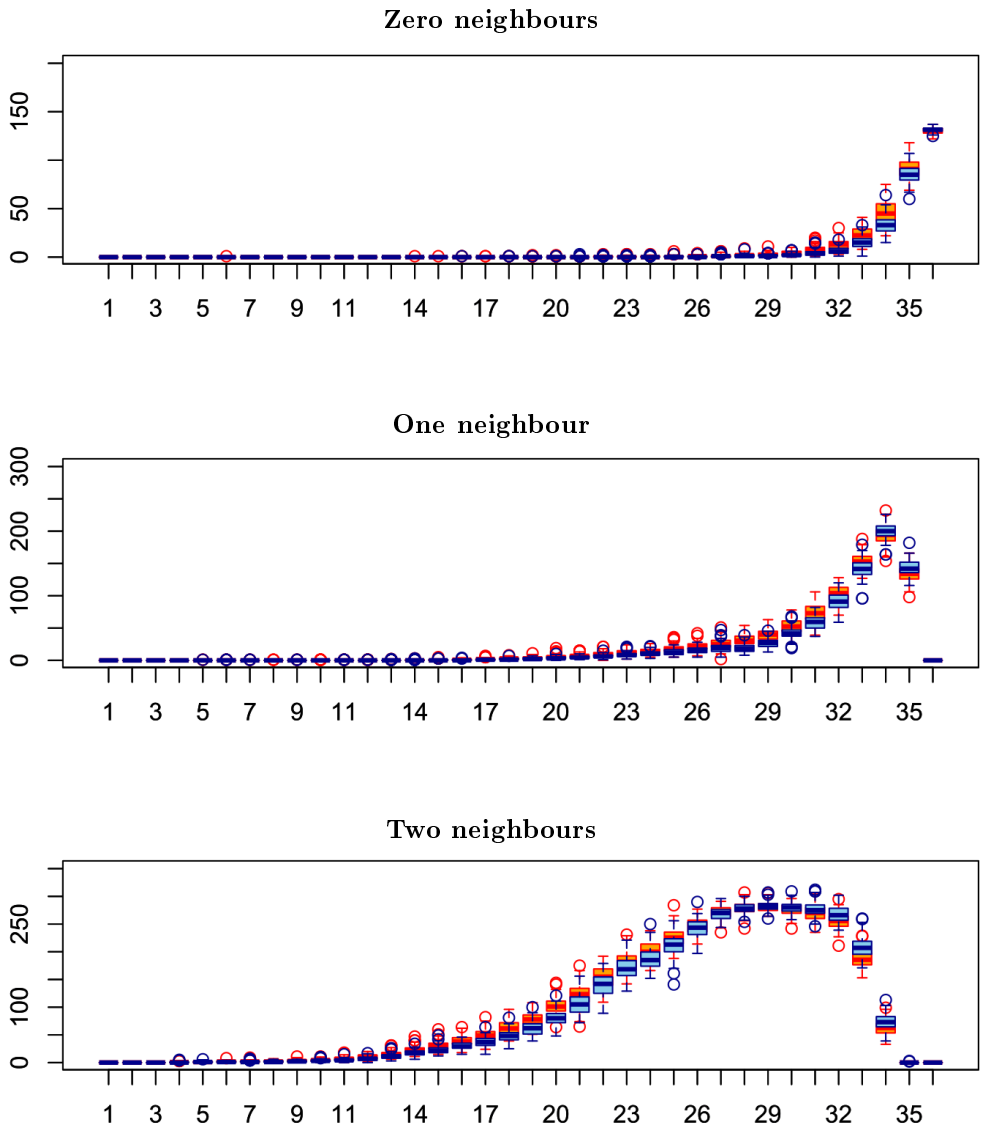


Figure 3.10: Results from stochastic simulations, relative to *Out* expressing cells. Each plot contains the distribution of the number of cells containing 0, 1 or 2 neighbours in the same state as themselves. Each box plot denotes the distribution of the number of neighbours expressing *Out* for a given *In* threshold and asynchronism (100 simulations). The σ-asynchronism has a role in reducing the number of cells with few same-state neighbours, when *In* thresholds are permissive to higher numbers of same-state neighbours. Although not statistically significative, the differences are consistent.

3.5 Pattern formation in growing tissues

3.5.1 Biological context

One of many challenges in developmental biology is to understand how growth affects tissue and organ morphogenesis. The uniformity of structures across organisms is due to the existence of extremely sensitive surveillance mechanisms acting at several scales, coordinating growth and shape.

How this coordination ensures developmental robustness remains a mystery. Tackling this issue requires the consideration of multiple concurrent processes at different levels. Most approaches resolve around particular subsets of properties associated phenotypic outcomes (*e.g.* cell shapes and consequent geometric packing of tissues).

Cells receive signals from their surrounding, and interpret them as a stimulus or inhibition for their division. Cell division contributes to additional readjustments of the population configuration and consequent effects on cells, in a tightly controlled feedback mechanism (Wartlick *et al.*, 2011).

Environmental cues largely contribute to tissue expansion and shape and can be separated into two main groups: morphogenic signals are diffusible components of heterogeneous distribution that rule the architecture of tissue growth during early developmental stages. Here, cell division inhibition leads to smaller organs and tissues and altered phenotypes. For example in vertebrate limb pattern formation, inhibition of cell division leads to digital loss (Alberch & Gale, 1983). Nutrient dependent growth affects the size of organs, while maintaining the morphology unaffected, as demonstrated in nutrient starved organisms (Hietakangas & Cohen, 2009).

Similarly, physical constraints such as compression and mechanical feedbacks between cells have been shown to directly correlate with expansion capabilities of tissues (Streichan *et al.*, 2014). These can be decomposed in local effects of cell division orientation, cell competition, shape and rearrangement (Lecuit & Le Goff, 2007).

Provided we can abstract the full complexity of this process, it is possible to address some of the mechanisms involved in growth of developing tissues. Several

computational approaches have been developed for the purpose. These usually assess the problem in terms of cell shapes and volumes (physical descriptions), and how the emergence of new cells (cell divisions) affect highly packed epithelia.

Vertex models consider cells to be irregular polygons defined by edges and vertexes. These completely define cell attributes (such as area, shape and number of neighbours). Changes to their properties are largely due to motion of the vertexes. Motion is regulated by a set of underlying continuous energy functions acting on the junctions of neighbour cells, and dependent on the mentioned attributes. The evolution of these systems is therefore a relaxation to mechanical equilibrium through changes of connectivity according to cell rearrangements. Incorporation of proliferation generally assumes that cell divisions occur stochastically rather than including a detailed description of the cell cycle (Fletcher *et al.*, 2014). Other possibilities include proliferation and death as result of cell area thresholds: above a given area, cells divide, and bellow a specified size, they die. Applications of this framework include cell sorting and tissue engulfment (Brodland, 2015), tissue size control and elongation (Farhadifar *et al.* (2007), Mao *et al.* (2011)), and other morphological patterns.

Some constructions allow for the definition of chemical feedbacks and morphogen transport among the cells. These consider small sets of components, their distribution and effect through partial differential equations, depicting simple regulatory mechanisms associated to cell patterning events (Fletcher *et al.* (2014), Kachalo *et al.* (2015)).

Cellular Potts models constitute a discrete framework to account for mesoscopic cell descriptions, first proposed by Graner & Glazier (1992). As in cellular automata, this framework relies on discrete arrays, but instead of mapping each cell to a position, the biological entities are modelled by a set of adjacent points or "pixels". The dynamics changes of cell shapes occur through Monte Carlo simulations determining whether cell pixels acquire the state of one of the neighbours, representing small deformations in cell membranes. In the original definition, cells have no internal structure (as in vertex models). Cells move or change their shapes by adding or excluding nodes according to a rule dependent on their current structure. Resulting behaviours visually resemble membrane fluctuations

and pseudopod formation (Voss-Böhme, 2012). As in vertex models, cell motility can be coupled with proliferation and death as stochastic or size dependent events. The coupling with reaction-diffusion equations and cell-type specification has also been considered.

Nonetheless, none of these approaches are easily transposable to our framework. The dynamics of conformation changes are made under the assumption that each step of a simulation is constant in elapsed time. For Logical Models, this is not true, as transitions between states may correspond to distinct time scales (and unknown rates).

Hence, we present an alternative approach to model tissue dynamics in a coarse grained framework where there are no explicit shape changes and tissue rearrangements must occur through cell displacements. The main assumptions for this approach regard cell topology. In order to maintain the same degree of abstraction, and the subsequent grid of hexagonal cells, we consider that all individuals are identical in size and shape, as well as number of neighbours. Moreover, mitotic events are instantaneously resolved by cells by immediate displacement, ensuring their topology is unchanged. In counterpart, thanks to the inclusion of Logical Models, instead of assuming the proliferative trigger is stochastic or associated to the topology, it can be associated to the Logical Model states and used for the study of pattern formation under constant remodelling conditions.

3.5.2 Regulation of cell division

Studies of cell cycle frequently rely on the detection of specific markers for cell cycle progression or arrest. In typical representations of cell cycle, a post-mitotic cell starts at the G_1 phase, during which it grows due to biosynthetic activity and cytoplasmic organelle duplication. After completion, the cell enters S phase where chromosome duplication occurs. At G_2 stage the cell prepares the required machinery for division, followed by M phase where the division occurs. Each of these stages is associated to cyclin-dependent kinase complexes that permit progression, and upon which regulatory mechanisms act (Vermeulen *et al.*, 2003).

Regulation occurs through checkpoints dependent on intra and extra-cellular cues. A checkpoint in G_1 acts in response of nutrient availability and growth fac-

tors presence, while additional controls further on are activated by DNA damage and spindle formation. In the G_1 checkpoint however, if the microenvironmental conditions are not favourable for cell cycle progression, cells become arrested in a G_0 state from which exit into cell cycle is only possible in the presence of growth factors. At the organism scale, this nutrient dependent arrest results in tissue atrophy due to starvation (Schafer, 1998).

Along the growth of tissues and developmental commitment, cells gradually lose proliferative capability, concomitant with increasing fate restriction. In the later case, cells can either become terminally differentiated and unable to further divide (*e.g.* peripheral nervous system cells, cardiac muscle cells), or retain proliferative abilities for the purpose of cell replacement, as in epithelial cells of most internal organs.

Network Boolean models have been used to analyse these regulatory circuits. These already include publications on the mammalian cell cycle (Fauré *et al.*, 2006), fission yeast cell cycle (Davidich & Bornholdt, 2008) or signalling interplay in neural specification or proliferation (and later epithelial fate) in *Drosophila* morphogenesis (Ghysen & Thomas, 2003), but their behaviour in a community of cells has still to be addressed.

Hence, because Logical Models are well adapted to model cell-fate decisions, the proliferative status of a cellular network can also be accounted for. This can be achieved by simply including a regulated node depicting cell division, or a module containing the regulatory components of such progression, whose state reflects the required markers of cell cycle progression.

Therefore, in our framework, we define a *proliferative pattern* that triggers a cell division. A pattern is a subspace of the state space of a Logical Model, in which a specified subset of the variables must respect a range of qualitative levels. Throughout the simulation steps, cells containing that Logical Model are tested for the proliferative pattern.

Upon division, an additional event in the grid is considered, in which a new cell appears next to the triggered cell. Two decisions must be made: the state of the sister cells, and the positioning of the new cell, as well as subsequent displacement of the neighbours. Our current implementation defines that the state of the sister cells is the same as the mother cell initial state. Moreover, if

the mother cell contains a perturbed node, the daughter cells will also carry that perturbation. Cell division is considered to always maintain planar polarity, and does not account for asymmetrical cell divisions nor the aforementioned intrinsic loss of proliferative abilities. This approach provides a simple base system to tackle the complexity Logical Models can impose.

3.5.3 Tissue rearrangements

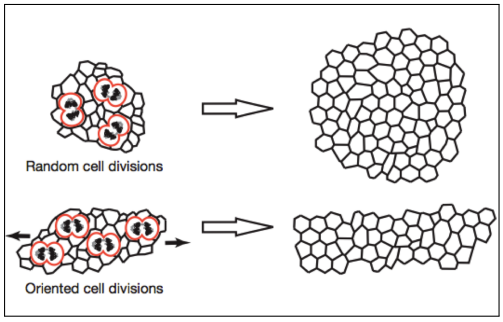


Figure 3.11: Morphogenesis in a tissue with uniform mitotic events can occur through random oriented cell divisions generating a concentric shape, or polarised cell divisions that lead to elongation (adapted from Lecuit & Le Goff (2007)).

The details on how individual cell behaviours affect morphogenesis and growth are still largely unknown. Hence, a possible starting point to computationally address this issue is to consider tissue rearrangements to occur through cell displacement. This rearrangement begins by the choice of cell division orientation. The orientation can either be random, leading to concentric tissue growth, or polarised due to some environmental cue, in which case the tissue elongates in accordance to it. In the later case, it has been shown for instance that the elongation of *Drosophila* embryonic epithelia is partially controlled by biased cell divisions along the proximal/distal axis (Lecuit & Le Goff, 2007).

The first step in this implementation is to extend the Cellular Automata framework to account for the existence of empty positions to where cells can divide and move in a growing tissue. The definition in Section 2.2.1 imposes that every cell in $\mathcal{L}$ was mapped to a state in $\mathcal{Q}$. We remove this restriction and consider that only a subset of cells have an attributed Logical Model (and hence an attributed state), while the remaining are empty. The cells containing a state of the Logical Model are representative of the biological entities, while others depict empty spaces or "slots". These empty positions still can be associated with virtual grid nodes accounting for constant environmental signals.

Here, we assume that changes to cell size and division are coupled in a single event. All cells are considered to be identical in size, shape, and adhesive properties. Because of the topological constraints, we need to rely on the very strong assumption that remodelling of the cells is fast enough to be considered instantaneous, *i.e.* cells immediately converge to the lowest energy conformation.

The rearrangement of the tissue is given by a displacement of cells, under the only restriction that the number of displacements should me minimised. The cells shift in accordance to the path towards the border that offers the minimum resistance. This approach was chosen in detriment of random path due to boundary conditions. Modellers may wish to avoid border continuity, by which choosing a random walk can result in traps (grid borders) from which it is not possible to leave. Therefore, the only requirement would be to develop a path search algorithm. The option of a *greedy algorithm* that searches exactly the shortest path was soon after rejected due to the fact that it generates hexagonally shaped tissues (an artefact resulting from CA constraints). If the proliferative events are spatially uniform, the tissue grows in the shape of an hexagon, due to the shortest path being a reflection of the grid topology.

Therefore, a new approach was considered. A measure was attributed to each cell, quantifying the number of its neighbours, and hence, the *compression* it suffers. The higher the compression, the "harder" it is to displace the cell. Several authors have considered this measure in order to study tissue size regulation, which can be translated into cytoskeletal mechanics, generating forces that are transmitted through cell-cell junctions (Heisenberg & Bellaïche (2013), Hufnagel *et al.* (2007)).

3.5.3.1 Compression level

Given a grid of cells $m \times n$, we define a Boolean matrix such that each element (i, j) $(i = 1, \ldots n,\ j = 1, \ldots m)$ indicates the presence or absence of a cell in that position of grid. If a position (i, j) is empty, the compression exerted in it is set to zero. Otherwise, if a position (i, j) is occupied, a vector $V_{i,j}$ is specified, whose size k is the maximum distance from which cells contribute for the compression level of the considered cell (i, j). The vector component at index d is the number

of cells (*i*.e. occupied positions) at distance d from the cell (i, j). Any cell at distance d is accounted for in that index if it has a *direct connection* with the cell i, j. A direct connection is a sequence of contiguous occupied positions at decreasing distances.

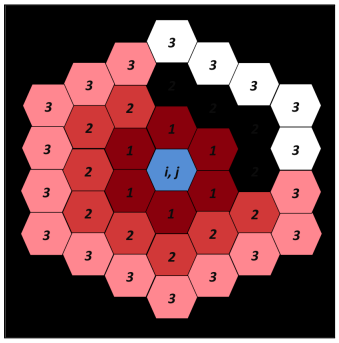


Figure 3.12: For a cell i, j (in blue), and a maximum distance of effect $k = 3$, red-coloured neighbours represent the contributors for the compression level. The lighter shade of red depicts a smaller contribution of those cells. White cells have a null contribution either because their distance to the cell i, j is too large, or because they are not directly connected to it. Black background represents unoccupied positions. The associated vector is $V_{i,j} = [6, 8, 13]$.

We propose an heuristic method relying on this vector of cell counts, to define the compression level $c(i, j)$ exerted on the cell (i, j). The higher the number of cells in the vector, the higher is the compression. Moreover, cells at larger distances contribute less to the compression than closer cells. Hence, the expression of the compression level combines the decreasing contribution of farther cells with the cell density effect. Note that for simplicity, to define this compression level, the grid is considered toroidal (*i*.e. no border conditions). If the grid borders are not continuous, the assumption is that cells at the border have the same number of neighbouring positions as the ones in the centre. The "missing" positions in the borders are assumed to be occupied positions.

For the first term of compression, we associate a weight (reflecting the contribution to the compression) that linearly decreases with the distance:

$$\forall_{d=1,\dots k+1}, w(d) = \frac{1-d}{k} + 1.$$

This weight linearly varies from 1 (for the closest neighbouring cells) to 0 (for $d = k + 1$).

Concerning the density, it obviously depends on the numbers of cells spanning the surrounding positions. Considering a cumulative contribution (*i*.e. linearly increasing with the number of cells), we have observed that the resulting shapes

of growing tissues have no biological relevance (star shaped artefacts). We thus propose to define that this contribution would follow a sigmoid curve.

Hence, without resorting to mechanical parameters, the sigmoid function below defines the contribution of the cells at a given distance to the compression level of the cell at position (i, j):

$$\forall_{d=1,...k}, f(d,i,j) = \frac{1}{e^{p\times(6\times d)-V_{i,j}(d)}},$$

where p is a parameter specifying the mid-point of the sigmoid, $6 \times d$ is the number of positions at distance d and $V_{i,j}(d)$ is the number of occupied positions at distance d from the cell i, j. The contribution of the number of cells can thus varie in the range $]0, 1]$.

Finally, the level of compression level $c(i, j)$ exerted on the cell at position i, j is defined as follows, combining the previous equations:

$$c(i,j) = \sum_{d=1}^{k} f(d,i,j) \times w(d).$$

Previous reports have shown that only closer neighbours (*i*.e. short ranges of distances) are responsible for the forces exerted on individual cells and for the maintenance of tissue morphologies when proliferation is not directed by a gradient (Escudero *et al.*, 2011). We will explore this observation further on, as we now possess an *in silico* structure to do so.

3.5.3.2 Algorithm

An algorithm retrieves the displacement path resulting from a cell division. The objective is to search for the path that goes down the slope of compression towards the tissue border.

It must take two possible cases into consideration (see Figure 3.13 for flowchart):

(a) The cells are heterogeneous in suffered compression. From the dividing cell, the displacement path is built from continuous cell selection under the condition that the sequential compression must decrease;

(b) The cells are homogeneous in suffered compression, a characteristic associated with increased distance from the tissue border. A concentric search of increasingly higher radius is preformed until at least one cell is different in compression. From that cell, one of several possible paths is randomly traced back to the original position. The remaining displacement path is built using case (a), as now the latest selected cell is in an heterogenous compression environment.

Upon cell division and placement of a new cell next to the triggered position, the remaining cells adjust by one position in the specified path. Unlike cell updates, cell division cannot occur synchronously due to the possibility of path conflicts. We consider a random order of divisions when concurrent cells are called for proliferation.

The algorithm is only called when there are still empty spaces in the lattice. If this condition is not satisfied, even though the cells may have calls for division, they will not divide.

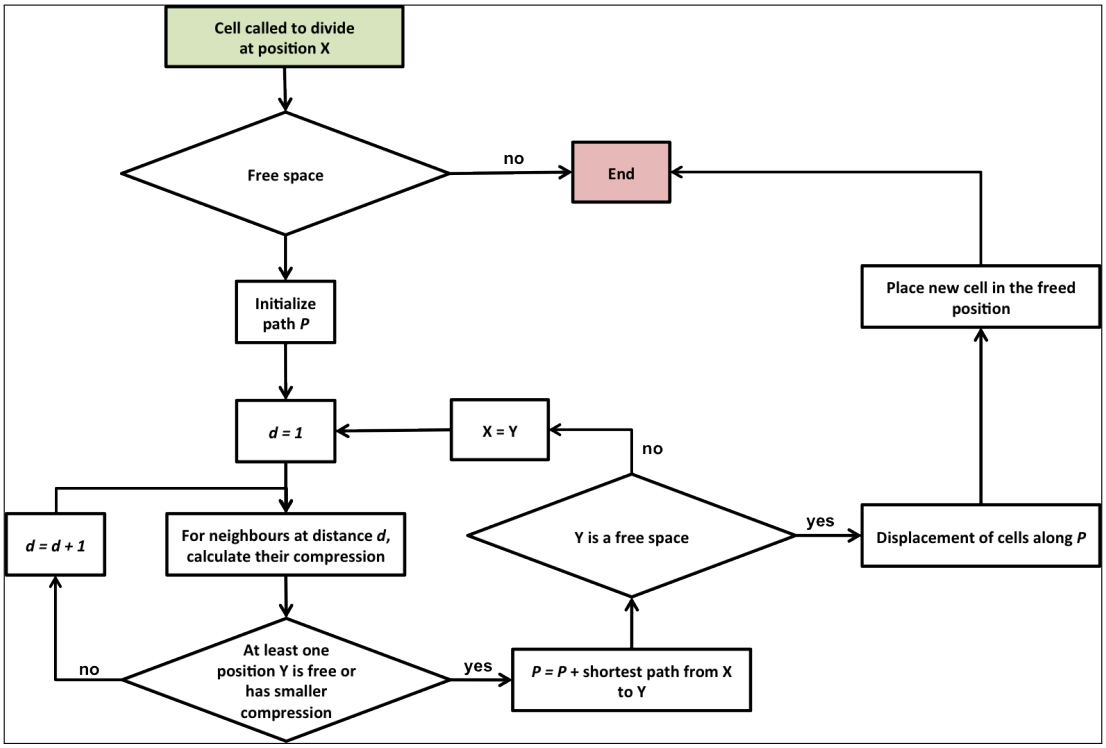


Figure 3.13: Algorithm flowchart for displacement path following a cell division.

3.5.4 Choice of compression parameters

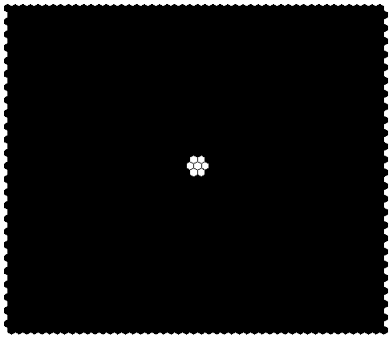


Figure 3.14: Initial cluster configuration used growth simulations, composed by 7 cells placed in a concentric shape. Black background represents the empty "slots".

If Logical Model states are not associated to the trigger for cell division, who it is set it to randomly occur in an α proportion of the cells, the distribution of dividing cells at each step is spatially uniform. Hence, starting from an initial condition with a concentric cluster of equivalent cells as presented in Figure 3.14, it is expected that the shape of a 2D growing epithelia is approximately concentric. As it has been mentioned, *greedy algorithms* always lead to hexagonal shapes. Our choice of parameter values for k and p should allow avoiding this artefact, while accounting for a mechanistic interpretation of cell displacement.

As expected (Table 3.1), short range effects od neighbours ($k = 3$) leads to homogeneous tissue growth in uniform divisions. Furthermore, compression is only increased when the number of neighbours approximates the maximum number ($p = 0.9$), *i.e*, the displacement is highly impaired when the number of empty "slots" is reduced.

If the midpoint proportion p is too low, the tissue shape will be similar to the one obtained with a greedy algorithm, regardless of the distance range considered. Curiously, for large distance ranges and a midpoint proportion of 0.5, the tissue growth becomes highly irregular, forming random branches. This occurs because the cells selected for displacement are favoured if they have less than half of their neighbouring positions occupied by cells.

For higher k and p, the tissue growth follows concentric branched morphologies. This is due to the higher sensitivity of the compression measure when accounting for larger radius.

Hence, the parameters were fixed at $p = 0.9$ and $k = 3$ as the ones that better approximate implicit mechanical feedbacks in homogeneous tissue growth.

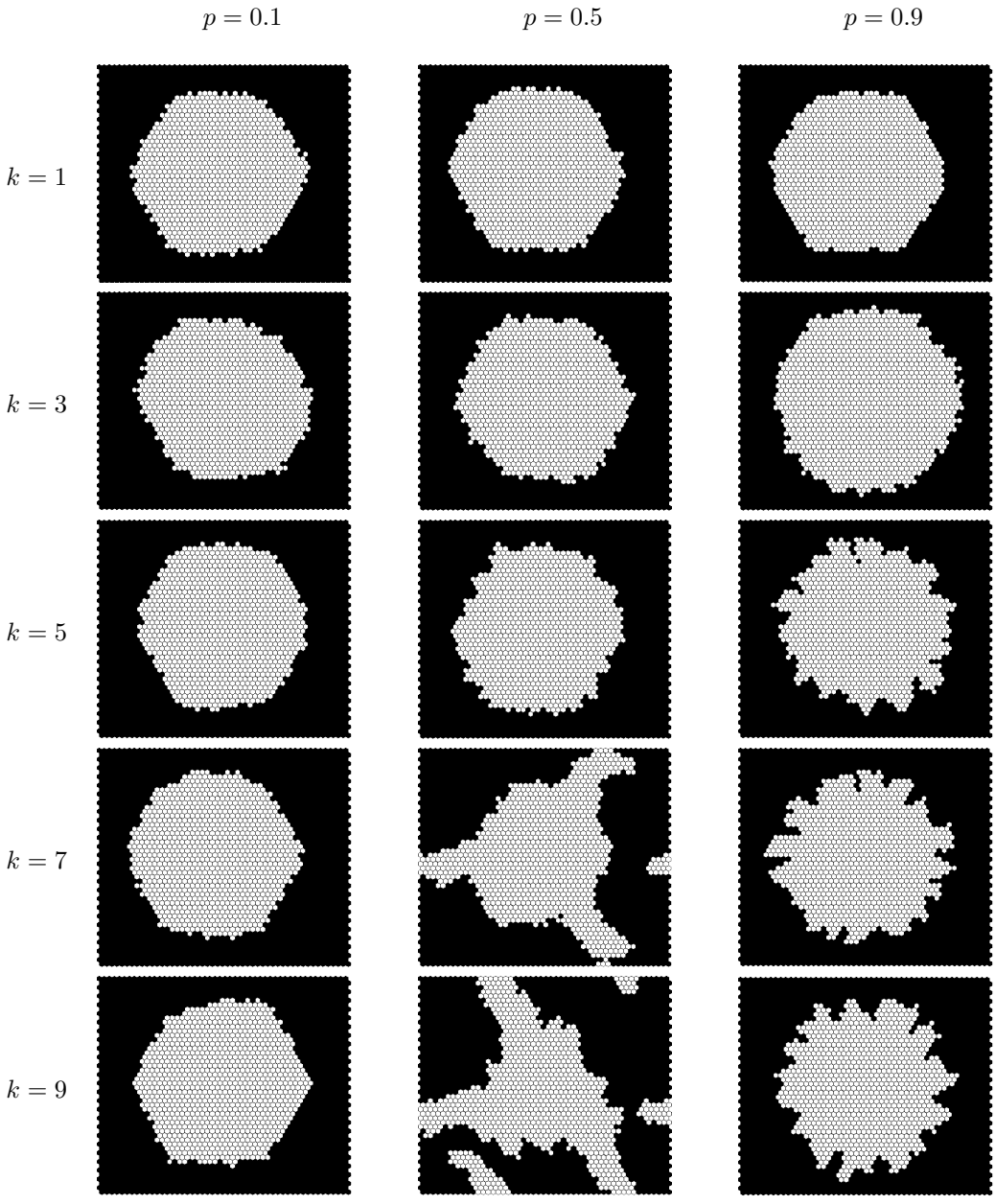


Table 3.1: Growth patterns resulting from changes to the parameters of compression in the cells. The simulations were performed in a 50×50 grid with toroidal conditions, and cells stoped proliferating after a population of 1000 individuals was attained. The initial conditions were the same for all cases, starting from a concentric cluster of 7 cells as in Figure 3.14.

3.5.5 Differential cell division in tissue morphogenesis

Cell cycle regulation through cell communication can lead to growing patterns that differ from the ones obtain through simple mechanical compression. This growth patterning will be dependent on the Logical Models. Furthermore, although mechanical compression plays an important role, it is not the sole contributor of tissue shaping. Other regulatory mechanisms include physical constraints inhibiting cell division, nutrient depletion resulting from local overpopulation and heterogeneous expression of growth factors.

In particular, branching morphologies in epithelial networks are well-studied developmental processes, ubiquitously identifiable across organs and organisms. Classical case studies include *Drosophila* tracheal system and other animal organs such as the lung, the kidney and the mammary gland (Ochoa-Espinosa & Affolter, 2012). These have been correlated with several regulatory mechanisms, one of which is the differential growth (Varner & Nelson, 2014).

Hence, a possible transposition of these properties is to consider that Logical Model inputs can also be regulated by a parameter depicting cell density, adding a new layer of environmental interactions that include mechanotransduction. As in integration functions, it can be defined as a cardinality constraint over the positions in a neighbourhood range. But here, the constraints are over the number of cells in the neighbouring positions.

An example of these implications can be illustrated using the toy model. A cell cycle trigger will simply corresponds to having the In component activated. The Out component acts as an inhibitor of the cell cycle. It is activeted by a threshold number of occupied positions found in a neighbouring range $[1, d_{max}]$. The tracking of mitotic activity can be followed by visualisation of cells expressing In. It is expected that the mechanical feedback provided by cell density c has some effect on the shape of the growing tissue, as mitotic events will no longer be uniform in time and space. As before, we fix d_{max} to three possible values (1, 3 and 6), to which the associated n_{max} are respectively 6, 18, and 126. The simulations were performed until a population of 1000 individuals formed or previous stabilisation occurred.

Our simulations in Table 3.2 recover the aforementioned relationship between differential growth and branching morphologies. Green cells are cells for which cell cycle is not inhibited, while red cells are unable to divide. For specific values of the c threshold, branching morphologies arise in the epithelia, and stabilise on their own (total number of cells is stable and smaller than the number of positions in the grid). Branch length and width vary in accordance with the neighbouring distance considered, and the threshold c affecting *Out* expression. Moreover, it is clear that there is a differential growth, depicted by the presence of green (proliferating) cells only at the tip of the branches.

For smaller thresholds, the inhibition of proliferation quickly settles in and the tissue stops growing. For larger levels of c, the morphology of the epithelium maintains its concentrical shape, but only cells closer to margins contribute for expansion.

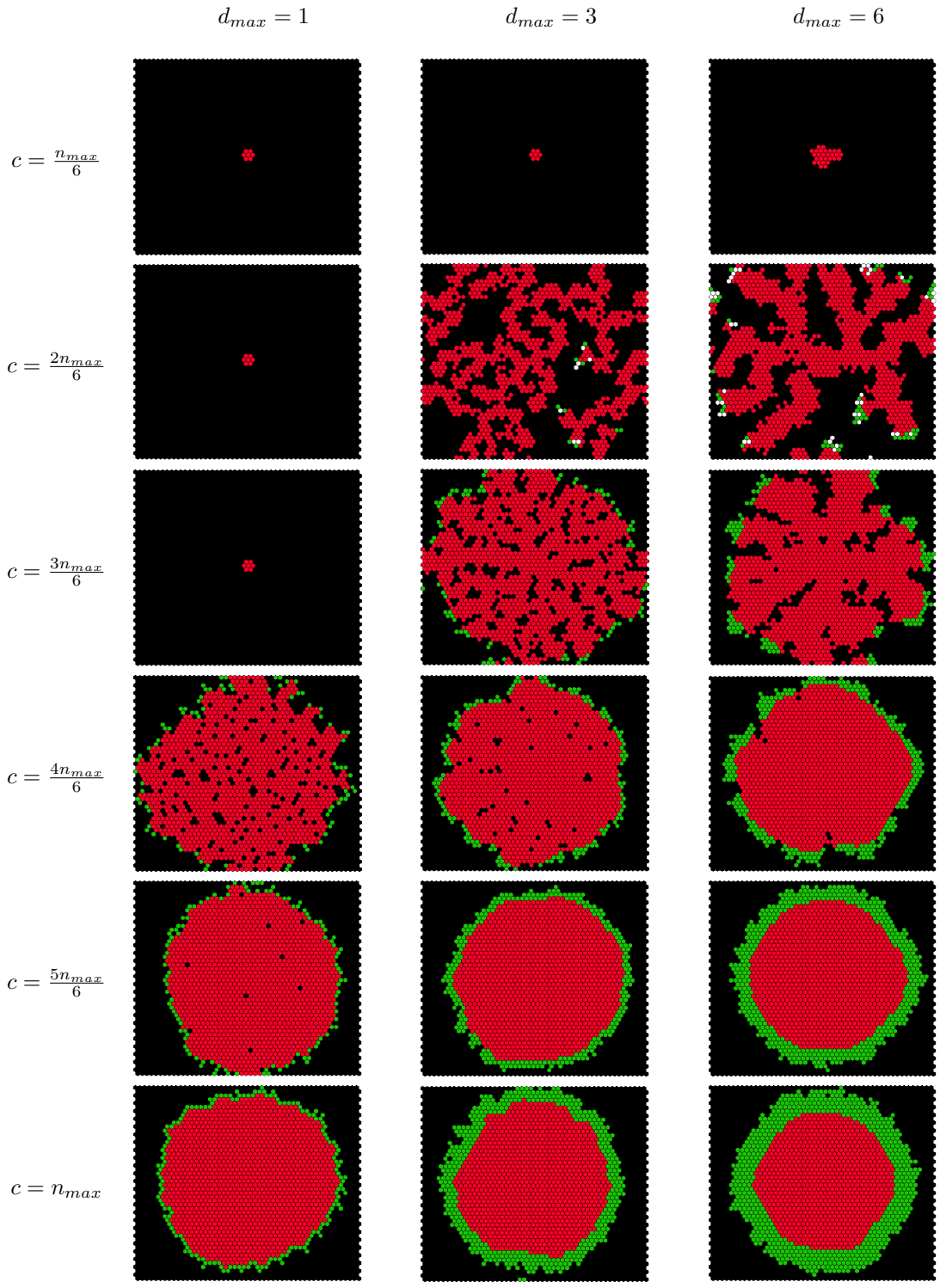


Table 3.2: Growth patterns resulting from regulation of integration inputs of Logical Models. The initial conditions used are the ones described in 3.14. Red (Out) cells are inhibited for proliferation, while green (In) cells will be triggered for division. c is the threshold number of occupied neighbours that result in activation of the Out component. For low c the cells are unable to proliferate.

Chapter 4

Case study: *Diptera* mechanoreceptor patterning

This chapter describes the case study built to further explore the functionalities of the established modelling framework as well as the resulting version of the tool EpiLog.

4.1 System description

Most invertebrates have sense organs distributed on their body surfaces. In particular, adult Diptera organisms usually have a large number of sensory bristles, whose positions and numbers vary between species. The diptera mechanoreceptors are components of the peripheral nervous system whose formation is spatially and temporally restricted, as their precursors are singled out of columnar epithelial cells (which form the adult epidermis) during the third instar larvae development (Usui & Kimura, 1992).

Upon commitment, these precursors, also called Sensory Mother Cells (SMC) undergo two asymmetrical divisions to produce four cells that will differentiate into a neurone, a sheath cell, a socket cell and a shaft cell, the main components of a sensory bristle (Usui & Kimura, 1992).

Two types of bristles can be distinguished in fly bodies: the macrochaetae -large bristles - are less numerous and have strictly invariant and reproducible positions for some Diptera species, making them useful for taxonomic purposes

(Furman & Bukharina, 2008); the small bristles in the notum of the fly organisms are more numerous and form more relaxed patterns that differ amongst clades. The small bristle pattern can range from random to fully aligned stripes of mechanoreceptors, along the anterior-posterior axis of the organism; in these stripes, time of appearance or location of the SMC seem to be fairly unrestricted (Simpson *et al.*, 1999), simply ensuring non-adjacency between SMC.

Similium variegatum *Calliphora erythrocephala* *Tryphera succinata*

Figure 4.1: The microchaete (black dots) and macrochaete (red dots) patterns in several *Diptera* species. The anterior-posterior alignment and the distance between bristles constitute the major differences between them. Adapted from Simpson *et al.* (1999).

The mechanisms underlying such rigorous developmental patterns in Sensory Mother Cell selection have been extensively studied in *Drosophila* since 1954, when Stern (1955) experimentally demonstrated that the positioning of the large bristles in the fly's notum was the result of a predetermined invisible pattern able to induce neural competence.

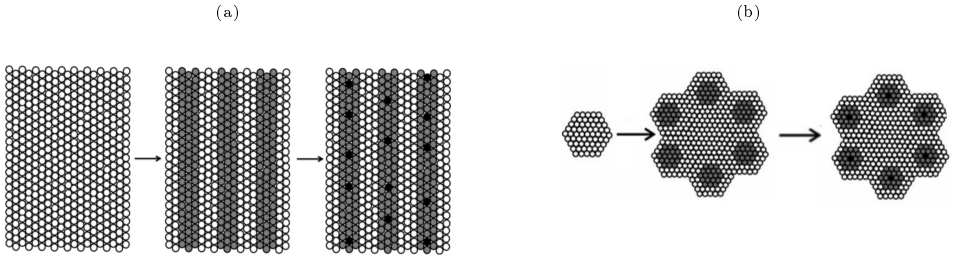


Figure 4.2: Steps in micro-and macrochaete SMC selection. **(a)** The epidermal tissue forms stripes of competence domains (grey cells) in which SMC (black cells) arise spaced from each other. **(b)** A small pool of cells proliferate, during which competence domains of proneural expression start to form at precise positions. Near the centre of the clusters, a cell becomes committed to neural fate, and it will later give rise to one large bristle (adapted from Furman & Bukharina (2008)).

The system responsible for responding to these prepattern components is centred on the *achaete-scute* (*as-c*) gene complex, commonly called proneural gene complex. The complex contains the sequences for four bHLH transcription units, as well as a set of individual enhancers for each unit. The complex products are transcription factors that positively regulate neuroblast fate (Gómez-Skarmeta *et al.* (2003), Campos-Ortega (1994)).

Mutations on individual enhancer sequences result in suppression of location-specific macrochaetae and disappearance of patches of microchaetae, while gain-of-function experiments over *as-c* result in ectopic external sense organs. These results suggest that, although the expression of individual transcription units depends on different prepatterning conditions, they all possess the same intrinsic activity as selector genes for neural development (Bertrand *et al.*, 2002). Furthermore, they also suggest that the underlying regulation of the SMC specification has overlapping features between both types of chaetes.

As shown in Figure 4.2, SMC selection is a three-stage process of complex cellular signalling and intracellular feedback responses. Our goal is to recover some of the details that lead to the formation of the patterns of SMC distribution, and to assess how within each group of prepatterning conditions the specification of the SMC occurs. The objective is to recover a general mechanism that could be extrapolated to other patterning events in mono-layered epithelia.

4.1.1 Formation of competence domains

A prepattern is defined by a group of transcription factors that together are able to induce *as-c* expression. A prepattern determines the domains of expression of proneural genes both for macrochaete SMC (in the form of clustered domains) and microchaete SMC (stripe domains) (Calleja *et al.*, 2002). Prepatterns are different in their composition of transcription factors. Therefore, in the same organism, frequently no two domains of expression react the same way to applied perturbations (*e.g.* gene knock-downs).

Moreover, the regulation of the prepattern domain formation is distinct in micro- and macrochaetae (Rudel & Sommer, 2003). This observation is based on the fact that large bristle precursor cells are positioned in a more restricted

manner than small bristles SMC. Moreover, mutational responses are largely different between both types of bristle precursors, and the formation of clusters for macrochaete SMC specification are highly sensitive to mutations.

Few proteins have been shown to directly activate the transcription of the *as-c* gene complex. These include the product of the *pannier* gene and the members of the *iroquois* complex (Leyns *et al.* (1996), Gómez-Skarmeta *et al.* (2003)). The expression of these genes is not uniform across the tissue. However, together, they cover the entire fly notum. It is typically in areas of overlapping expression that *as-c* expression is induced. In those domains a pool of equipotent cells forms. The cells in the pools of *as-c* expression are epidermal, but have the competence to become neuroblasts.

4.1.2 Lateral cooperation and SMC commitment

After the formation of the pool of competent cells, these become distinguishable from the remaining epithelial cells due to as-c expression. Nevertheless, selection of a cell for neural commitment depends on its ability to reach a threshold level of *AS-C* protein expression (Furman & Bukharina, 2008).

Proneural cells in the centre of the clusters (or respective midline of stripes) express relatively higher levels of AS-C before the SMC selection. Reports show that this is due to the activation of the Epithermal Growth Factor Receptor (EGFR) and signal transduction by Ras/Raf/MAP kinase cascade, which enhances the expression of *as-c* genes (Culi *et al.*, 2001).

The signal strength resulting from EGFR positively correlates with expression of proneural genes in surrounding cells, incrementing *as-c* expression in the centre of the competence groups and causing a smoothening out of its distribution (zur Lage *et al.* (2016), Culi *et al.* (2001), Ghysen & Thomas (2003)). This suggests *as-c* expression induces the production of a signal that is sensed by surrounding cells through EGFR. Cells in the centre are therefore able to reach the required threshold levels of AS-C that lead to SMC commitment through lateral cooperation (see Figure 4.3 for interaction scheme).

The sensitivity of proneural cells to EGFR-signalling differs in macrochaete and microchaete development, as small bristles competent cells present a much

higher resistance to EGFR perturbations (Culi *et al.*, 2001).

These results are in agreement with the fact that a minimum number of cells in the competence domains is required to form the bristle pattern (Furman & Bukharina (2008), Parks *et al.* (1997)). For small mechanoreceptor stripes, the required width is usually reduced to about three to five cells. In the case of macrochaete clusters however, these must contain about 20-30 cells, and are more prone to fail the SMC commitment upon perturbation of EGFR signalling.

4.1.3 Lateral inhibition and SMC isolation

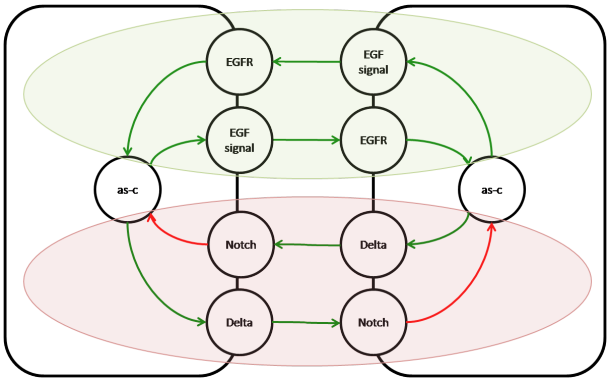


Figure 4.3: Simplified scheme of the communication between two proneural cells. Green arrows denote positive regulations, and red arrows repressions. The nodes in the green background provide the necessary potentiation of as-c expression, while the components in the red background are responsible for lateral inhibition, ensuring only one cell reaches the highest levels of as-c expression.

In wild type conditions, two SMC never arise adjacent to each other. This results from lateral inhibition mediated by Delta-Notch signalling. The main components of this system are the transmembrane receptor Notch, its ligand Delta, and the intracellular target genes of the active component of Notch, Nidc (Notch intra-cellular domain) (Furman & Bukharina, 2008).

Delta-Notch signalling pathway is a widely conserved inter-cellular communication system that plays a key role in cell-fate specification via local interactions between cells. Activation of the intra-cellular signal-transduction causes the inhibition of the neuroblast fate by a transcriptional repression of *as-c* expression (Artavanis-Tsakonas *et al.*, 1999). Inhibited cells retain an epidermal fate.

Notch activation and release of its intra-cellular domain is mediated by a Delta-like ligand that binds to the membrane receptor. Delta is a transmembrane component, acting through juxtacrine communication. In our system, the

production of Delta-like ligand is directly regulated by the proneural genes. These bind to Delta's promoter region and activate its transcription (Parks *et al.*, 1997).

SMC express both the receptor Notch and the Delta ligand. An isolated SMC prevents the surrounding cells from neuroblast fate by Delta binding. Mutual inhibition occurs when two adjacent cells reach the threshold expression of *as-c*. In this case, they both compete to maintain a neural fate by sending repressing signals to each other. Eventually, one of the cells is able to prevent its neighbours from attaining a neural fate and becomes the single precursor cell.

Delta-promoted neural repression is usually short-ranged. However, SMC in the notum of the fly are able to extend filopodia that mediate long-range inhibition. The minimum distance that separates two SMC depends on the range of the inhibition field. Therefore, the higher this range, the higher is the distance separating two SMC.

In large bristle formation, each cluster produces one SMC. For small bristles sensory specification, each stripe can contain a variable number of SMC in various positions of the stripe. The relative positioning of these cells simply ensures the inhibition range is respected.

4.1.4 SMC selection in proliferating tissues

The regulation of mitotic activity plays a central in growth control, commitment and differentiation. The formation of the sensory bristles in *Drosophila* is compartmentalised both in time and space, and they both depend on this regulation. Large bristles SMC are specified during the late instal larva, while the tissue is exponentially growing. Patterns of small bristles precursor cells are only defined after growth arrests.

Literature on cluster formation is inconsistent. Some reviews suggest that the prepattern results from a positional morphogenic signal that does is not affected by changes that occur on the growing monolayered tissue, and hence, the cluster formation is a simple consequence of as-c induction in cells located in those positions (Ready *et al.*, 1976). If so, proliferation *per se* does not play a role in cluster formation, but contributes to the spreading of the pool of epithelial cells able to respond to such stimuli. Other authors propose that the prepatterning signal

acts on a smaller pre-existent pool of cells, which maintains that competence and proliferate until correct cluster size has been attained.

The spatial distribution of mitotic activity is not uniform in the growing notum region.The proneural clusters are mitotically quiescent groups of cells (Usui & Kimura, 1992), and *as-c* expression correlates with the arrest of mitotic activity in those regions. After the SMC is singled out of this population, the remaining cluster decreases proneural expression and resumes its mitotic activity.

String (*str*) is a known molecular marker that controls the transition from G_2 to M phase, inducing mitotic activity. Proneural genes negatively regulate the expression of *str*. Moreover, *str* leads to transcriptional repression of *as-c* genes through *E-spl*, involved in the same pathway as *Notch* signalling. This network forms a negative feedback that switches cell fates between epidermal and neural precursor, to precisely coordinate tissue growth and patterning (Nègre *et al.* (2003), Ghysen & Thomas (2003)).

Stripe domains however only occur after proliferation stops. The regulation of stripe patterns for small bristles possibly relies on some feedback mechanism. Inter-stripes in the notum of the fly are know to have high level of expression of antagonist transcription factors, that cannot coexist with *as-c* expression. A Turing mechanism could thus explain the stripe patterns, relying on these antagonistic genes.

4.2 Logical Model definition

We propose two distinct Logical Models in Figure 4.4 to account for cell fate decisions in large and small bristle patterning variants, taking into account the information gathered in Section 4.1. In these networks, the central regulatory contour is similar in effects over the *as-c* expression. However, cell communication events that lead to SMC selection should be different, because the spatial and temporal specifications of competence domains differ.

Hence, our Logical Models encode the crucial events that occur within a cell, causing it to take a neural fate or maintain epithelial properties.

Two questions can be assessed through these Logical Models and their incorporation in an epithelial context:

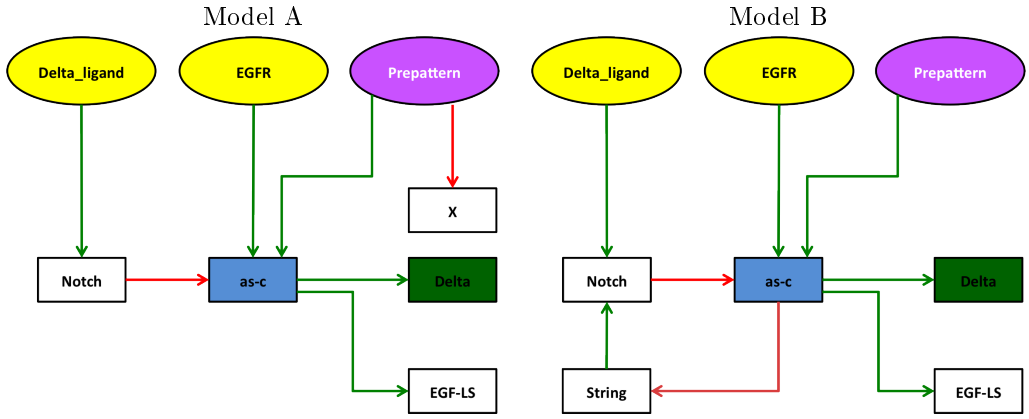
- Given an initial prepattern domain, is it possible to recover SMC selection mechanisms;

- Considering the methods implemented in EpiLog, is it possible to obtain the prepattern configuration associated to each type of bristles (clusters *vs* stripes).

Both models include three inputs that regulate the expression levels of proneural genes: the Prepattern, responsible for the initial induction of proneural expression, the EGFR pathway, promotor of as-c expression, and the Delta_ligand, repressor of proneural expression. EGF-LS is the EGF-like signal induced by as-c, serving as an activator of EGFR expression in neighbouring cells.

Model A of small bristles does not account for proliferation because this patterning occurs when the tissue has stopped growing. Moreover, it considers an additional component, X, antagonist for prepattern gene expression. We propose that at this stage of development the pattern of expression of competent domains could result from a cell non-autonomous feedback circuit between Prepattern node in one cell and the putative diffusible component X from surrounding cells. For large bristles, Model B considers an additional component, String, antagonist to as-c expression. This component defines the trigger for proliferation.

The stable states of Logical Models A and B can be categorised into three cell fates: component as-c at zero is associated to epidermal fate; as-c level at one is associated to an epidermal fate, still competent for a neural fate (this denotes a cell of a proneural cluster or stripe); as-c at its maximum level is associated to a neural fate, for which the cell becomes a bristle cell. Neural fate is also characterised by the presence of Delta expression (see Table 4.1).



Component	Level	Logical Function
Notch	1	Delta_ligand
as-c	2	Prepattern & EGFR & !Nicd
	1	Prepattern & Nicd | !EGFR
Delta	1	as-c:2
EGF-LS	1	as-c
X	1	!Prepattern

Component	Level	Logical Function
Notch	1	String | Delta_ligand
as-c	2	Prepattern & EGFR & !Nicd
	1	Prepattern & Nicd | !EGFR
Delta	1	as-c:2
EGF-LS	1	as-c
String	1	!as-c

Figure 4.4: The Logical Models for SMC fate decision with respect to the central regulator *achaete-scute* complex. Ellipses are inputs. Purple ellipse is the inductive prepattern. Red arrows represent negative regulations, while green arrows correspond to activations. Model A: Small bristle regulatory network, comprising an unknown putative component X that helps establish stripe patterns of competent cells. Model B: Large bristle regulatory network, comprising a String node to account for cell proliferation.

Input values			Stable state				
Prepattern	EGFR	Delta_ligand	Nicd	as-c	Delta	EGF-LS	Phenotype
0	*	*	0	0	0	0	Epidermal
1	0	0	0	1	0	1	Competent epidermal
1	*	1	1	1	0	1	Competent epidermal
1	1	0	0	2	1	1	SMC

Table 4.1: The stable states obtained from all combinations of input values. The stable states of the shown components are the same for both Models A and B. Furthermore, in Model A the epidermal stable state contains the putative component X activated. Similarly, in Model B the epithelial stable state is associated to String expression.

4.3 Epithelial definitions

Lateral cooperation is driven by one cell expressing EGF-LS, and its positive effect in a cell receiving this signal through the input component EGFR. The sensitivity of precursor cells to this signal varies between competence domains. For small bristle patterning, cells are less sensitive to altered expression of the signal, hence the differences in the integration functions presented in Table 4.2. Lateral inhibition drives the effect of one neighbouring Delta signal inducing Notch pathway in a cell.

For large bristles patterning (Model B), mitotic events in the epithelium are given by the activation of String, leading to tissue expansion and clonal propagation of cells that express the prepatten genes.

For small bristles, instead of clonal propagation, the inducer of the expression of prepattern genes is the diffusible component X.

		Model A Small bristles	Model B Large bristles
Communication	EGFR=1	EGF-LS(1, 6, 6, 1)	EGF-LS(1, 18, 18, 1:2)
	Delta_ligand=1	Delta(1, 1, _, 1:3)	Delta(1, 1, _, 1:3)
	Prepattern=1	X(1, 70, _, 1:6)	None
Proliferation trigger		None	String=1

Table 4.2: Integration functions for the Models A and B in the context of an epithelium (see Section 3 for the syntax of these functions).

4.4 Results

4.4.1 Model A: Formation of small bristle patterns

In the pattern of small bristles the competence fields are long stripes where equidistant SMC arise randomly. If a regulatory mechanism ensures that the competence fields are stripes of prepattern signals, random locations of these SMC would follow the stripe orientation. The systematic analysis of the role of the integration functions in inducing patterning (Section 3.4) allows us to postulate that combination of spots inside stripes could be obtained through an

interaction ensuring stripe formation, and a subsequent mechanism that is activated within the stripes and generates spot patterns in that domain. The spot patterns are merely isolated SMCs. The prepatterning is therefore expected to be the result of extensive communication by a population of initially naive cells, under the effects of α and σ-asynchronism.

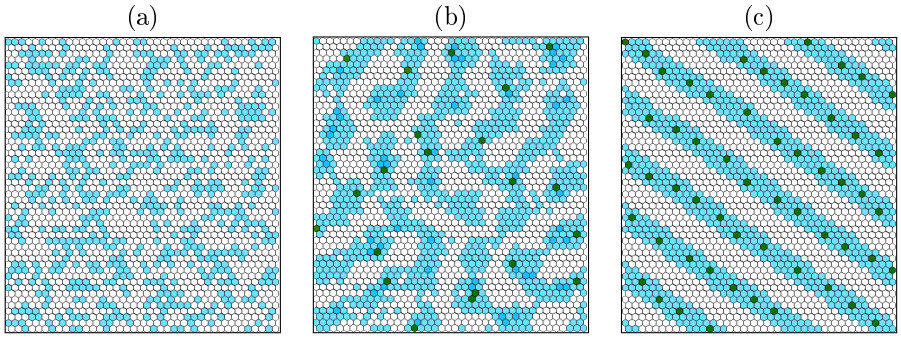


Figure 4.5: The evolution of the dynamical system comprises three qualitative stages. The colour scheme is the same as for components in Model A, Figure 4.4. Blue cells express as-c, and dark green cells express both as-c and Delta, as committed SMCs. The simulation was performed under $\alpha = 0.2$, $\sigma_X = 0.2$. **(a)** In the early stages the distribution of *as-c* expressing cells is random (and so is the subsequent prepattern expression). **(b)** Halfway the simulation, the prepattern domains start to aggregate and form irregular stripes, some of them wide enough to momentaneously allow cells to reach the highest levels of *as-c* expression. **c** When the stable state occurs the stripe domains are evenly separated, and contain specified SMC in irregular positions.

.

SMC commitment is given by stable expressions of as-c at level 2 and Delta. The remaining as-c expressing cells are part of the competence domain, and contribute to the maintenance of the SMC by lateral cooperation through EGFR pathway. Although not accounted for in our model, the cells will eventually decrease the expression of proneural genes and proceed with epithelial fate.

The orientation of the stripes can follow any of the hexagonal cells axis because there is no specification of anterior-posterior domains. Forcing an anterior-posterior orientation (*i.e.* removing horizontal boundary conditions) the pattern aligns vertically, but the stable state remain qualitatively the same. The stripe width varies between 3 cells and 5 cells, like specified in the wild type case.

4.4.2 Model B: Formation of large bristle patterns

The analysis of the pattern formation of large bristles was separated into two phases, firstly to ensure the specification of a single SMC in the clusters, and secondly to check that the required size and shape of the clusters could be reproduced.

Hence, we first considered a set of simulations in which the prepattern expressing domains are already formed. According to Furman & Bukharina (2008), normal SMC specification occurs in concentric clusters of about 20-30 cells.

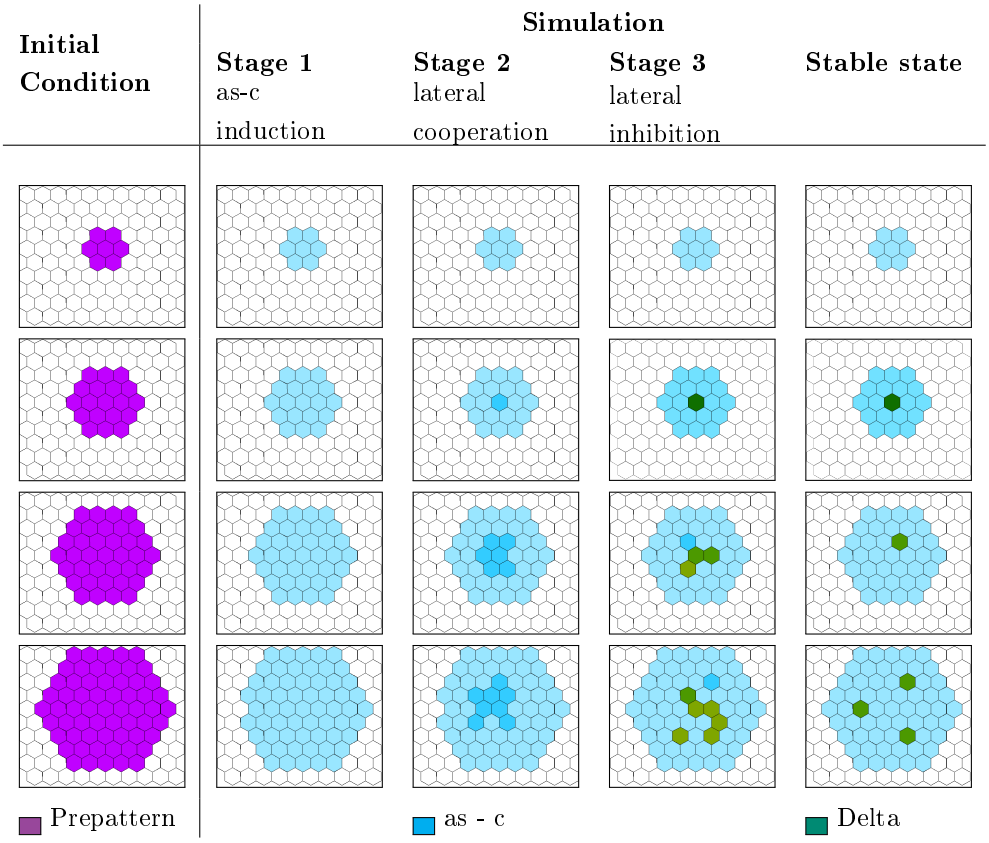


Figure 4.6: The evolution of the large bristle SMC specification for several cluster sizes (rows). The clusters are defined by the presence of the Prepattern component in the initial condition (first column). In the remaining steps of the simulation, the component was not selected for visualisation because it remains constant. The simulations were performed under α-asynchronism at 0.2. Neural cells contain the highest levels of *as-c* expression (darker blue). Their isolation is ensured by the expression of the Delta component in green. For small clusters the communication is not sufficient to induce cooperation and subsequent lateral inhibition.

The dynamics of cluster formation play an important role in SMC determination. Figure 4.6 demonstrates how the width of the cluster of prepattern expression affects the existence and total number of SMC. For small clusters, the number of cells is not sufficient to induce lateral cooperation, and none of the cells can surpass the required as-c threshold to commit to neural fate. If the clusters are too large, several cells receive stimuli to become SMC. Those that overcome the threshold begin expressing Delta in order to inhibit surrounding cells from doing the same. However, this repression mechanism is not strong enough to reach further away cells that have committed to neural fate as well. Cluster sizes that allow for the occurrence of exactly one SMC, are robust to variation of the number of cells between 19 and 37 concentric elements.

Upon validation of wild-type patterns of SMC formation within each cluster, we now couple the process with proliferation. We present two different hypotheses for cluster formation in growing tissues:

- Hypothesis 1: The competence domains form through clonal propagation of a smaller pool of prepattern expressing cells (Furman & Bukharina, 2008);

- Hypothesis 2: The competence domains are induced by a static field of a morphogen defining the prepattern. Tissue growth follows the morphogenic field that defines the clusters.

In the first hypothesis, we consider a small pool of cells where one only already expresses prepattern genes. Tissue growth is induced by the String component. Hence, it would be expected that the cell expressing prepattern genes would propagate concentrically, as the remaining tissue does. The growth should eventually lead to the formation of a sufficiently large cluster to allow for SMC specification.

We were not able to recover any form of cluster when accounting for growth, as sister cells would always disperse due to the displacements caused by division of other cells. Hence, for patterns of large bristles, we recovered within-cluster properties, but were unable to simulate the mechanisms leading to cluster growth.

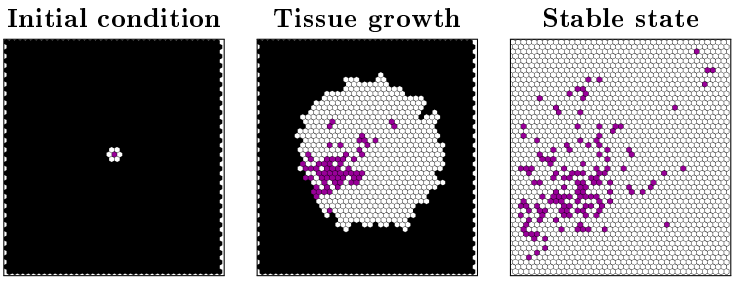


Figure 4.7: Simulation of the large bristle cluster formation from a smaller pool of cells, in a 40×50 grid with toroidal conditions. Cells associated with prepattern expression (purple) do not form a cluster, but rather disperse until full tissue size is reached.

Moreover, this observation remains valid under the second hypothesis. It is possible to recover SMC specification but only after the tissue stops growing. Otherwise, in the intermediate steps, cells that are positioned in the morphogenic field become competent, but are quickly removed from the prepattern location by displacement, as in the previous scenario.

Developmental studies show that SMC specification occurs during tissue growth (and not after) and this case study shows that our method to account for proliferating epithelia has its limitations (see Section 5.3).

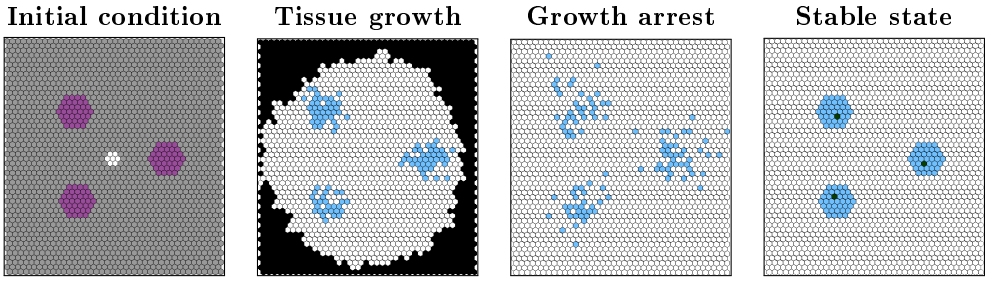


Figure 4.8: Simulation of the large bristle cluster formation due to a morphogenic field, in a 40×50 grid with toroidal conditions. In the initial condition grey positions are empty spaces. There, three fields of prepattern signals are applied. The tissue grows and as-c expression is induced, but the cells are unable to maintain cohesion. Only when the grid dimensions force mitosis to stop do the clusters (and subsequent SMCs) form, as now there are no displacements occurring.

Chapter 5

Discussion

5.1 Overview of the framework extensions

The current implementations of the framework and subsequent exploratory analyses of its capabilities provided insights into the extensions developed within this work. Although discrete systems have been considered inappropriate to model tissue morphogenesis (Macklin *et al.*, 2010), several models of epithelial growth can be transposed to logical abstraction.

The defined measure of compression is able to tackle CA constraints in growing tissues, that otherwise attain hexagonal-like shapes regardless of the initial conditions. Moreover, it is possible to specify a feedback response to mechanical stress regulating cell cycle arrest through signalling mechanisms (Hufnagel *et al.* (2007), Varner & Nelson (2014), Heisenberg & Bellaïche (2013)). Our measure of population density by a simple cell count, which ultimately affects cell division, is a coarse abstraction of the several mechanisms known to be involved. The overall effect of these on each cell can all be correlated to the number of surrounding cells (Varner & Nelson, 2014).

Modelling of density dependent growth in concentrically growing tissues (*i. e.* the later cases of Figure 3.2) has been related to organ size regulation. For example, morphogenic signals such as Decapentaplegic (Dpp) play important roles in mitosis induction and morphogenesis of epithelia. Their effects depend on inductive threshold concentrations. Hufnagel *et al.* (2007) proposed that the

concentration gradient is unaffected by monolayered tissue growth, and size regulation by growth arrest depends on an antagonistic interplay between Dpp and mechanical compression exerted over cells. Hence, cells should arrest when too far away from the source of the morphogen and under low compression, or when they are near the morphogen source, but the tissue is overgrown and mechanical stress transduction inhibits mitosis.

Our proposed method is able to implement these observations, with the advantage that the transduction mechanisms operating under mechanical stress and morphogenic signals can be extensively specified and coupled with patterning signals in a Logical Model involving all these mechanisms.

5.2 Microchaete patterns

Several authors have proposed epithelial models for small bristle determination. Cohen *et al.* (2010) considered a 2-state automaton where each cell is epithelial or neural. Applying cardinality constraints to define cell communication, coupled with appropriate signalling noise that provides the required stochasticity, the authors recover well aligned stripes of competent cells. Here, there was no definition of prepatterning conditions, or how they regulate SMC specification (Cell specification occurs directly through lateral inhibition). Similarly, Kachalo *et al.* (2015) consider a vertex model in which fields of Delta and Notch expression are applied and upon which the behaviour of the cells is defined through coupled diffusion equations. In the stripe domains it is possible to form aligned SMCs.

The authors chose simple models that allowed for the analysis of a more central focus of study (respectively the role of signalling noise, and the development of a vertex model tool), and provided few mechanistic definitions. No explicit mechanism of stripe formation was accounted for.

Our approach considers the stripe domains to be self-organising, and the formation of the SMCs to be dependent on a set of components associated to biological entities. Hence, the model is able to recover important properties of the system. Our apparently simple approach, containing two interlaced patterning mechanisms dependent on cellular networks, demonstrates that such an abstract framework can account for complex patterns in life.

5.3 Macrochaete patterns

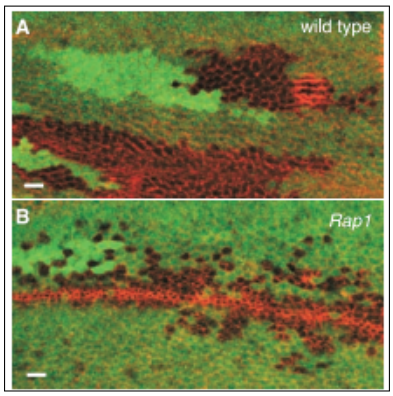


Figure 5.1: Cells in *Drosophila* wing disc are separated into two populations, one of which expresses GFP. In normal adhesion, the clusters of similar cells are in contact (upper image). In mutant organisms the cells scatter and no clusters form (bottom image). Figure adapted from Knox & Brown (2002).

Our model was unable to recover the cluster formation in *Diptera* macrochaete. The differential adhesion between cells, which is mediated by E- and N-cadherins was not taken into account in this model, nor in the proposed consideration of cell division in the modelling framework. Upon cell division, there is no remodelling of cell contacts between the two daughter cells. This is the major contributor to why clones remain compact in growing tissues. Defects in the distribution of E-cadherin after cell division lead to a disruption of cell contacts and to cell scattering (Lecuit & Le Goff, 2007). Defects in this even distribution after mitosis cause disruption of cell contacts and cell scattering. Our patterns of clonal cells are remarkably similar to those presented in Figure 5.1. Hence, the current implementation of proliferation is not suited for patterning models that imply clonal propagation.

5.4 Further work

There is room for improvement of our framework to support proliferation in network-dependent cellular automata. In order to further explore patterning in growing tissues we need to address the aforementioned observations on cell scattering. To this end, it will be necessary to account for cell lineage. By adding a new layer to the framework and keeping track of the relationship between neighbouring cells (whether they are sister cells or not, for example) it should be possible to account for adhesive properties. Coupling this method with compression measures will provide a displacement function that recovers both morphogenetic properties (qualitative changes to the shapes of the tissues), and some internal

dynamics of cell motions in the displacements. More precisely, it may be possible to account for cell displacements that form a rotation around sister cells, in order to maintain the aggregates.

Moreover, the system used to define proliferation can be easily adapted to account for cell death (apoptosis). Because programmed cell death is also a highly regulated process, this event can be defined as a state pattern of the Logical Models allocated to the cells. Here, cells may or main not work their way into readjusting to the empty space. In cavitation for example, apoptosis has a function in forming hollow structures, and therefore no readjustment occurs to fill the increasingly larger gap. Moreover, apoptosis is a necessary process that ensures tissue homeostasis and cleansing of aged or altered cells in organs (Abud, 2004). As a consequence, apoptosis can occur alone without causing reconfiguration in the neighbouring cells (morphogenesis through apoptosis), or it may be feasible to consider dynamically homeostatic situations, where cell proliferation is locally induced by signalling from dying cells, ensuring the maintenance of tissue cohesion.

Finally, along with the compression measures, it would be useful to account for oriented cell divisions (*i.e.* along a gradient), and specification of sister cell states after mitosis, to account for asymmetric cell divisions or gradual loss of ability to proliferate.

Chapter 6

Concluding remarks

Theories provide the necessary bridge between experimental observations, and related interpretations from scientists. If it is possible to formalise those interpretations, it is possible to assess if they can explain the data.

In model development, there is an inherent trade-off between abstraction and realism. With the increase of computational power and development of specific methods and tools in this area, the trade-off is becoming evermore skewed, and underlying assumptions in theoretical work are becoming relaxed.

Logical modelling assumes that the modelled systems can be handled through a discrete abstraction and they do not require the precise knowledge on kinetic parameters. When adequate, this is thus a powerful abstraction to qualitatively describe large-scale, long-term dynamics, and to *e.g.* efficiently predict interventions to alter unwanted behaviours.

Composition of cellular networks in an epithelial context provides increasing insights into mechanisms involved and possible interventions to retrieve correct developmental processes. This approach performs remarkably well in recovering patterning events that occur in tissues, despite the absence of explicit temporal dynamics. In addition, the framework is able no only to recover pattern formation mediated by activation-inhibition mechanisms (responsible for the well-known Turing patterns), it has also been proven to be well-suited in defining of heterogeneous conditions and formation of complex structures.

The work developed during this thesis provided an overview of several limitations in the cellular automata modelling of tissue patterning, and led to the

definition and implementation of several extensions of biological relevance that can undermine these limitations. The incorporation of epithelial dynamics of growth demonstrated how a coarse-grained approach can still account for complex patterning events, with the advantage that it is possible to observe cell-specific differences and distinct contributions of subpopulations to pattern formation. The comparison with experimental results is direct, since the data is generally obtained by detection of molecular markers. If these are defined in the network models, they can be individually observed in the epithelial context.

Nonetheless, considering the extent of information in the literature in need for a proper definition and annotation of a model, logical modelling is a complex task. It is common to detect inconsistencies leading to opposite conclusions, and these frequently result from insufficient knowledge on the core regulation of the process under study. Moreover, the boundaries of a Logical Model are also subjected to a necessary cutoff due to incomplete knowledge on cross-talks between the involved pathways. This has to be done ensuring that the model fairly reproduces the behaviour of key components.

In our framework, this issue is aggravated by the fact that studies on growth and morphogenesis cannot easily perform single-cell tracking. Our key purpose is to provide an abstraction that connects single cell behaviour to the tissue patterning and morphogenesis.

The fact that we cannot recover the process of cluster growth under the implemented method for cell proliferation (in Chapter 4) is a consequence of the lack of connection between the available information. Nonetheless, the results comprising cell scattering provided insights into the basal requirements we need to take into account when addressing the issue of epithelial dynamics, specifically in cellular automata designed for compact epithelia modelling.

Overcoming this drawback, the framework can become a powerful asset in experimental work.

References

Abou-Jaoudé, W., Monteiro, P.T., Naldi, A., Grandclaudon, M., Soumelis, V., Chaouiya, C. & Thieffry, D. (2015). Model checking to assess t-helper cell plasticity. *Frontiers in Bioengineering and Biotechnology*, **2**. 10

Abou-Jaoudé, W., Traynard, P., Monteiro, P.T., Saez Rodriguez, J., Helikar, T., Thieffry, D. & Chaouiya, C. (2016). Logical modeling and dynamical analysis of cellular networks. *Frontiers in Genetics*, **7**. 2, 7, 9, 10

Abud, H. (2004). Shaping developing tissues by apoptosis. *Cell Death & Differentiation*, **11**, 797–799. 64

Alberch, P. & Gale, E.A. (1983). Size dependence during the development of the amphibian foot. colchicine-induced digital loss and reduction. *Development*, **76**, 177–197. 33

Artavanis-Tsakonas, S., Rand, M.D. & Lake, R.J. (1999). Notch signaling: Cell fate control and signal integration in development. *Science*, **284**, 770–776. 51

Bahi, J.M. & Contassot-Vivier, S. (2002). Stability of fully asynchronous discrete-time discrete-state dynamic networks. *IEEE Transactions on Neural Networks*, **13**, 1353–1363. 9

Bertrand, N., Castro, D.S. & Guillemot, F. (2002). Proneural genes and the specification of neural cell types. *Nature Reviews Neuroscience*, **3**, 517–530. 49

REFERENCES

Bouré, O., Fatès, N. & Chevrier, V. (2011). *Robustness of Cellular Automata in the Light of Asynchronous Information Transmission*, 52–63. Springer Berlin Heidelberg, Berlin, Heidelberg. 11, 26, 28, 29

Brodland, G.W. (2015). How computational models can help unlock biological systems. *Seminars in Cell and Developmental Biology*, **47**, 62 – 73. 34

Calleja, M., Renaud, O., Usui, K., Pistillo, D., Morata, G. & Simpson, P. (2002). How to pattern an epithelium: lessons from achaete-scute regulation on the notum of drosophila. *Gene*, **292**, 1 – 12. 49

Campos-Ortega, J. (1994). Genetic mechanisms of early neurogenesis in drosophila melanogaster. *Journal of Physiology-Paris*, **88**, 111 – 122. 49

Chaves, M., Albert, R. & Sontag, E.D. (2005). Robustness and fragility of boolean models for genetic regulatory networks. *Journal of theoretical biology*, **235**, 431–449. 9

Cohen, M., Baum, B. & Miodownik, M. (2010). The importance of structured noise in the generation of self-organizing tissue patterns through contact-mediated cell–cell signalling. *Journal of The Royal Society Interface*. 26, 27, 28, 29, 62

Culi, J., Martin-Blanco, E. & Modolell, J. (2001). The egf receptor and n signalling pathways act antagonistically in drosophila mesothorax bristle patterning. *Development*, **128**, 299–308. 50, 51

Davidich, M.I. & Bornholdt, S. (2008). Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE*, **3**, 1672. 36

de Back, W., Zhou, J.X. & Brusch, L. (2012). On the role of lateral stabilization during early patterning in the pancreas. *Journal of The Royal Society Interface*, **10**. 14

Escudero, L.M., da F. Costa, L., Kicheva, A., Briscoe, J., Freeman, M. & Babu, M.M. (2011). Epithelial organisation revealed by a network of cellular contacts. *Nat Commun*, **2**, 526. 40

Farhadifar, R., Röper, J.C., Aigouy, B., Eaton, S. & Julicher, F. (2007). The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, **17**, 2095 – 2104. 34

Fauré, A., Naldi, A., Chaouiya, C. & Thieffry, D. (2006). Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics*, **22**, e124–e131. 36

Fauré, A., Vreede, B.M., Sucena, É. & Chaouiya, C. (2014). A discrete model of drosophila eggshell patterning reveals cell-autonomous and juxtacrine effects. *PLoS Comput Biol*, **10**, e1003527. 22, 23

Fletcher, A.G., Osterfield, M., Baker, R.E. & Shvartsman, S.Y. (2014). Vertex models of epithelial morphogenesis. *Biophysical Journal*, **106**, 2291 – 2304. 34

Furman, D. & Bukharina, T. (2008). How drosophila melanogaster forms its mechanoreceptors. *Current Genomics*, **9**, 312–323. 48, 50, 51, 58, 59

Ghysen, A. & Thomas, R. (2003). The formation of sense organs in drosophila: A logical approach. *BioEssays*, **25**, 802–807. 36, 50, 53

Gómez-Skarmeta, J.L., Campuzano, S. & Modolell, J. (2003). Half a century of neural prepatterning: the story of a few bristles and many genes. *Nature Reviews Neuroscience*, **4**, 587–598. 49, 50

Gramss, T., Bornholdt, S., Gross, M., Mitchell, M. & Pellizzari, T. (2005). *Computation in Cellular Automata: A Selected Review*, 95–140. Wiley-VCH Verlag GmbH and Co. KGaA. 13

Graner, F.m.c. & Glazier, J.A. (1992). Simulation of biological cell sorting using a two-dimensional extended potts model. *Phys. Rev. Lett.*, **69**, 2013–2016. 34

Heisenberg, C.P. & Bellaïche, Y. (2013). Forces in tissue morphogenesis and patterning. *Cell*, **153**, 948–962. 38, 61

Hietakangas, V. & Cohen, S.M. (2009). Regulation of tissue growth through nutrient sensing. *Annual Review of Genetics*, **43**, 389–410. 33

Hufnagel, L., Teleman, A.A., Rouault, H., Cohen, S.M. & Shraiman, B.I. (2007). On the mechanism of wing size determination in fly development. *Proceedings of the National Academy of Sciences*, **104**, 3835–3840. 38, 61

Jernvall, J., Newman, S.A. *et al.* (2003). Mechanisms of pattern formation in development and evolution. *Development*, **130**, 2027–2037. 3

Kachalo, S., Naveed, H., Cao, Y., Zhao, J. & Liang, J. (2015). Mechanical model of geometric cell and topological algorithm for cell dynamics from single-cell to formation of monolayered tissues with pattern. *PLoS ONE*, **10**. 34, 62

Kim, S.H., Park, S., Mostov, K., Debnath, J. & Hunt, C.A. (2009). Computational investigation of epithelial cell dynamic phenotype in vitro. *Theoretical Biology and Medical Modelling*, **6**, 1–21. 13

Knox, A.L. & Brown, N.H. (2002). Rap1 gtpase regulation of adherens junction positioning and cell adhesion. *Science*, **295**, 1285–1288. 63

Le Novère, N. (2015). Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics*, **16**, 146–158. 2

Lecuit, T. & Le Goff, L. (2007). Orchestrating size and shape during morphogenesis. *Nature*, **450**, 189 –192. 33, 37, 63

Leyns, L., Gómez-Skarmeta, J.L. & Dambly-Chaudiere, C. (1996). iroquois: a prepattern gene that controls the formation of bristles on the thorax of drosophila. *Mechanisms of Development*, **59**, 63 – 72. 50

Macklin, P., Edgerton, M.E., Lowengrub, J.S. & Cristini, V. (2010). Discrete cell modeling. In *V. Cristini and J.S. Lowengrub, Multiscale Modeling of Cancer: An Integrated Experimental and Mathematical Modeling Approach*, chap. 6, 88–122, Cambridge University Press, Cambridge, UK, (invited author: P. Macklin). 61

Maini, P.K., Woolley, T.E., Baker, R.E., Gaffney, E.A. & Lee, S.S. (2012). Turing's model for biological pattern formation and the robustness problem. *Interface Focus*, **2**, 487–496. 25, 27

Mao, Y., Tournier, A.L., Bates, P.A., Gale, J.E., Tapon, N. & Thompson, B.J. (2011). Planar polarization of the atypical myosin dachs orients cell divisions in drosophila. *Genes and Development*, **25**, 131–136. 34

Nègre, N., Ghysen, A. & Martinez, A.M. (2003). Mitotic g2-arrest is required for neural cell fate determination in drosophila. *Mechanisms of development*, **120**, 253–265. 53

Ochoa-Espinosa, A. & Affolter, M. (2012). Branching morphogenesis: From cells to organs and back. *Cold Spring Harb Perspect Biol*, **4**. 44

Parks, A.L., Huppert, S.S. & Muskavitch, M.A. (1997). The dynamics of neurogenic signalling underlying bristle development in drosophila melanogaster. *Mechanisms of Development*, **63**, 61 – 74. 51, 52

Přibyl, M., Muratov, C.B. & Shvartsman, S.Y. (2016). Discrete models of autocrine cell communication in epithelial layers. *Biophysical Journal*, **84**, 3624–3635. 18

Ready, D.F., Hanson, T.E. & Benzer, S. (1976). Development of the drosophila retina, a neurocrystalline lattice. *Developmental biology*, **53**, 217–240. 52

Regnault, D., Schabanel, N. & Thierry, É. (2008). *On the Analysis of "Simple" 2D Stochastic Cellular Automata*, 452–463. Springer Berlin Heidelberg, Berlin, Heidelberg. 13

Rudel, D. & Sommer, R.J. (2003). The evolution of developmental mechanisms. *Developmental biology*, **264**, 15–37. 49

Schafer, K.A. (1998). The cell cycle: A review. *Veterinary Pathology Online*, **35**, 461–478. 36

Shou, W., Bergstrom, C., Chakraborty, A. & Skinner, F. (2015). Theory, models and biology. *eLife*. 1

Simpson, P., Woehl, R. & Usui, K. (1999). The development and evolution of bristle patterns in diptera. *Development*, **126**, 1349–1364. 48

Stern, C. (1955). Two or three bristles. *Science in progress*, 41. 48

Streichan, S.J., Hoerner, C.R., Schneidt, T., Holzer, D. & Hufnagel, L. (2014). Spatial constraints control cell proliferation in tissues. *Proceedings of the National Academy of Sciences*, **111**, 5586–5591. 33

Thomas, R. (1973). Boolean formalization of genetic control circuits. *Journal of theoretical biology*, **42**, 563–585. 2

Thomas, R. (1991). Regulatory networks seen as asynchronous automata: A logical description. *Journal of Theoretical Biology*, **153**, 1 – 23. 2

Turing, A.M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **237**, 37–72. 25

Usui, K. & Kimura, K.I. (1992). Sensory mother cells are selected from among mitotically quiescent cluster of cells in the wing disc of drosophila. *Development*, **116**, 601–610. 47, 53

Varner, V.D. & Nelson, C.M. (2014). Cellular and physical mechanisms of branching morphogenesis. *Development*, **141**, 2750–2759. 44, 61

Vermeulen, K., Van Bockstaele, D.R. & Berneman, Z.N. (2003). The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Proliferation*, **36**, 131–149. 35

Voss-Böhme, A. (2012). Multi-scale modeling in morphogenesis: A critical analysis of the cellular potts model. *PLoS ONE*, **7**, 1–14. 35

Wartlick, O., Mumcu, P., Julicher, F. & Gonzalez-Gaitan, M. (2011). Understanding morphogenetic growth control - lessons from flies. *Nat Rev Mol Cell Biol*, **12**, 594–604. 33

White, S.H., del Rey, A.M. & Sánchez, G.R. (2007). Modeling epidemics using cellular automata. *Applied Mathematics and Computation*, **186**, 193 – 202. 13

Zartman, J.J., Cheung, L.S., Niepielko, M.G., Bonini, C., Haley, B., Yakoby, N. & Shvartsman, S.Y. (2011). Pattern formation by a moving morphogen source. *Physical biology*, **8**, 045003. 22

zur Lage, P.I., Powell, L.M., Prentice, D.R., McLaughlin, P. & Jarman, A.P. (2016). Egf receptor signaling triggers recruitment of drosophila sense organ precursors by stimulating proneural gene autoregulation. *Developmental Cell*, **7**, 687 – 696. 50